

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP05/002066

International filing date: 26 February 2005 (26.02.2005)

Document type: Certified copy of priority document

Document details: Country/Office: EP

Number: 04360018.8

Filing date: 26 February 2004 (26.02.2004)

Date of receipt at the International Bureau: 08 April 2005 (08.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

Europäisches
PatentamtEuropean
Patent OfficeOffice européen
des brevets

26. 02. 2005

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

04360018.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk





Anmeldung Nr:
Application no.: 04360018.8
Demande no:

Anmeldetag:
Date of filing: 26.02.04
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Mixis France S.A.
166 Boulevard du Montparnasse
75014 Paris
FRANCE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Generation of recombinant genes in prokaryotic cells by using two
extrachromosomal elements

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State>Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/68

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI



Description

The present invention relates in general to in vivo methods for generating and detecting recombinant DNA sequences in prokaryotic cells, in particular bacteria, by using two different extrachromosomal elements and extrachromosomal elements, in particular plasmids that can be used for conducting the inventive methods. DNA sequences for which these methods are relevant include protein-encoding and non-coding sequences.

Evolution is a continuous process of genetic variation and phenotypic selection. The genetic diversity of a population can be amplified by creating new mutant combinations improving the performance of individuals within the population. The directed evolution of microorganisms has been accomplished traditionally through the process of classical strain improvement, by sequential random mutagenesis and screening.

Directed evolution has so far been used almost exclusively as a tool for engineering proteins. By mutation techniques such as site-directed mutagenesis, cassette mutagenesis, random mutagenesis, and error prone PCR variants of protein functions have been generated and the libraries thus produced have been screened for their ability to perform a specific function. Recursive application of this procedure has been used successfully for the modification of physical and catalytic properties of enzymes such as pH optima, thermostability, solvent stability, stereoselectivity, catalytic activity and substrate specificity as well as toxicity resistance mechanisms in bacteria and the host range and stability of viruses.

Traditional mutagenesis approaches for evolving new properties in enzymes have a number of limitations. These approaches are only applicable to genes or sequences that have been cloned and functionally characterized and that have a discrete function. Also, the 5 traditional mutagenesis approaches can only explore a very limited number of the total number of permutations, even for a single gene. However, under certain circumstances it might be necessary to modify not only one gene, but additional genes, in order to express a protein with new properties. Such additional genes can be for example 10 genes that cooperatively confer a single phenotype or genes that have a role in one or more cellular mechanisms such as transcription, translation, post-translational modifications, secretion or proteolytic degradation of a gene product. Attempting to individually optimize all of the genes having such function by traditional mutagenesis 15 approaches would be a virtually impossible task.

Most of the problems associated with conventional mutagenesis approaches can be overcome by recombination approaches which entail randomly recombining different sequences of functional genes, enabling the molecular mixing of naturally similar or randomly mutated genes. In comparison to conventional mutagenesis with recombination, the probability of obtaining mutants with improved phenotype is significantly higher. The main advantages of recombination approaches over conventional DNA manipulation technologies are in particular the experimental simplicity and the freedom from DNA sequence-imposed limitations. 20 25

Much of what is known about recombination processes has come from studies of simple, unicellular organisms such as bacteria. The study of recombination in such organisms has the advantage of

the ease of manipulation of DNA sequences and the possibility of studying specific recombination events induced synchronously in a large proportion of cells. Equally important is the growing conviction that the processes studied in microorganisms are identical or similar

- 5 in most respects to the ways in which mammalian cells, e.g. human cells, generate genetic diversity. Moreover, defining these mechanisms has taken on added importance in the quest to develop more efficient mechanisms of gene targeting and gene replacement in mammalian cells.
- 10 Although numerous different systems for effecting recombination in prokaryotic cells exists most of these do not allow an easy and reliable detection of newly recombined DNA sequences. Therefore, there is still in the art a demand for efficient prokaryotic test systems, which in particular allow a rapid and simple detection of recombinants and/or a selection of recombinants under selective pressure.
- 15

Therefore, the technical problem underlying the present invention is to provide improved methods and means for a simple and efficient generation of recombinant mosaic genes in prokaryotic cells, in particular for screening and detecting such recombinant sequences.

- 20 The present invention solves this underlying technical problem by providing a process for generating and detecting recombinant DNA sequences in prokaryotes comprising the steps of:
 - a) generating a first prokaryotic cell containing a recipient DNA molecule, which comprises a first DNA sequence to be recombined and which can autonomously replicate in the prokaryotic cell, and a donor DNA molecule, which comprises a second DNA sequence to be recombined and at least a first
- 25

- marker sequence encoding a gene product and which cannot autonomously replicate in the prokaryotic cell,
- 5 b) cultivating the first prokaryotic cell under selective conditions which only allow the growth and/or propagation of the cell if the gene product of the first marker sequence is expressed, and
- 10 c) isolating a second prokaryotic cell grown and/or propagated under selective conditions and containing a hybrid DNA molecule with the at least first marker sequence and a first and a second recombined DNA sequences due to recombination between the first and the second DNA sequences.

The present invention provides a prokaryotic system to screen for recombination events between at least two diverging, or heterologous, DNA sequences or recombination substrates *in vivo*. The inventive system allows the generation of new advantageous DNA sequences with improved properties in a fast and efficient way from at least two diverging DNA sequences by a process involving an *in vivo* exchange of DNA from two extrachromosomal elements containing the two DNA sequences to be recombined. The two extrachromosomal elements showing no homology in their nucleotide sequences are introduced into a prokaryotic host cell in the form of a recipient DNA molecule and a donor DNA molecule. The recipient molecule can autonomously replicate in the host, whereas the donor molecule does not have the ability to replicate. However, the donor molecule contains at least one unique protein-encoding marker sequence such as an antibiotic resistance marker or nutritional marker which is neither present in the genome of the host cell nor in the recipient mole-

cule. After the introduction of both extrachromosomal elements into the prokaryotic host cell the cell is cultivated under conditions which force the recombination between the two heterologous genes. These conditions can include for example a cultivation of the cell in the

5 presence of an antibiotic to which the cell is normally sensitive or a cultivation of the cell in a medium lacking an essential nutrient, which the cell cannot synthesize itself and which therefore has to be externally supplied. Under the cultivation conditions applied the cell can only grow and propagate, i.e. divide, if the gene product of the respective marker sequence is expressed. A prerequisite for the expression of the marker sequence is, however, that the non-replicative donor molecule and the replicative recipient molecule form a co-integrate which can replicate from the origin of the recipient molecule and thus ensures the maintenance of the marker sequence. The

10 formation of this co-integrate results from recombination between the two recombination substrates leading to the generation of new DNA molecules whose sequences differ from that of the parental DNA sequences. Host cells grown and propagated under the selective conditions applied therefore contain new recombinant DNA sequences.

15 The inventive process therefore provides an easy and quick selection system to identify recombinant DNA sequences. With the inventive process a large library of recombinant, mutated DNA sequences can be easily generated, and variants that have acquired a desired function can then be identified by using an appropriate selection or screening system.

20

25

The study of recombination processes in a simple, unicellular organism such as a prokaryote has the obvious advantage of the ease of manipulation of DNA sequences and the possibility of studying specific recombination events induced synchronously in a large propor-

tion of cells. Furthermore, over the last few decades a wealth of expertise has been accumulated both in the fermentation technology and the basic genetics of prokaryotic organisms. Another major advantage of prokaryotic host cells relates to their very short doubling times. Thus, by the use of prokaryotic host cells it is possible to conduct many rounds of cell division and thus many rounds of recombination and to create a plurality of new recombinant DNA sequences within a short time.

The inventive process can be conducted either in wild-type or mismatch repair-defective prokaryotic cells. The processes by which damaged DNA is repaired and the mechanisms of genetic recombination are intimately related; and it is known that the mismatch repair machinery has inhibitory effects on the recombination frequency between divergent sequences, i.e. homeologous recombination. Mutations of the mismatch repair system therefore greatly enhance the overall frequency of recombination events in prokaryotic cells. On the other hand, it is known that prokaryotic wild-type cells have a mismatch repair-dependent recombination mechanism, which is based on distantly spaced mismatches in two recombination substrates. Depending on the DNA sequences to be recombined, either wild-type or mismatch repair-defective prokaryotic cells can be used to obtain recombined sequences.

The inventive process for generating and detecting recombined DNA sequences has the advantage that greatly diverging DNA sequences can be recombined. It was unexpectedly found by the inventors that sequences with a high degree of overall divergence and which only share very short stretches of homology or identity can be recombined. An analysis of recombined sequences revealed that the long-

est stretches of identity in which recombination occurred comprise only 18-22 nucleotides. Most of the recombination events occurred in homologous stretches with a length of 10-15 nucleotides. In some cases recombination occurred in stretches with a length of only 4-5

5 nucleotides.

Therefore, the inventive process allows the recombination of DNA sequences derived from different prokaryotic species or different prokaryotic genera in order to create recombinant DNA sequences with advantageous features.

- 10 The inventive process for generating and detecting recombined DNA sequences has the advantage that more than two diverging sequences can be recombined, whereby the two diverging sequences to be recombined can be either pre-selected or non-selected sequences.
- 15 For example a variety of diverged DNA sequences up to a whole gene library can be inserted into the recipient as well as the donor DNA molecules. Afterwards the individual diverged DNA sequences can be recombined by chance with themselves. It is also possible to insert a first set of diverged DNA sequences up to a whole gene library only into the recipient DNA molecules and to insert a second set of diverged DNA sequences up to a whole gene library into the donor DNA molecules or vice versa and then to recombine the two sets by chance with each other. In both cases there is no selection as to which pair of diverged DNA sequences will be recombined.
- 20
- 25 However, it is also possible to recombine several, preferably pre-selected, diverging sequences in a stepwise manner, whereby in each step a selection is done as to which pair of diverged DNA se-

quences shall be recombined. If, for example, three diverging DNA sequences shall be recombined, than in the first step first prokaryotic cells are generated, whereby for example, a recipient DNA molecule with a first DNA sequence to be recombined and a donor DNA molecule with a second DNA sequence to be recombined are introduced into bacterial cells of a given species. These prokaryotic cells are cultivated under selective conditions such, that only cells are grown and propagated, in which the respective recipient and donor DNA molecules have formed a hybrid molecule enabling the expression of the marker sequence of the donor molecule and recombination between the two DNA sequences to be recombined. Thus second prokaryotic cells are obtained containing a hybrid molecule with a first and a second recombined DNA sequences due to recombination. The first and second recombined sequences are isolated and one of them is inserted for example into a recipient molecule, whereas the third DNA sequence to be recombined is inserted into the donor molecule. Then, in the next step, the recipient molecule comprising one of the recombined sequences and the donor molecule comprising the third DNA sequence to be recombined are introduced again in a prokaryotic host cell and subjected another round of recombination.

If, for example, four diverging DNA sequences shall be recombined, then in the first step two different sets of first prokaryotic cells are generated. For example, a first set of first prokaryotic cells can be generated by introducing a recipient DNA molecule with a first DNA sequence to be recombined and a donor DNA molecule with a second DNA sequence to be recombined into prokaryotic cells. Likewise a second set of first prokaryotic cells can be generated by introducing a recipient DNA molecule with a third DNA sequence to be re-

combined and a donor DNA molecule with a fourth DNA sequence to be recombined into cells of the same species. Each set of prokaryotic cells is cultivated under selective conditions to effect recombination. Thus a first set of prokaryotic cells is obtained containing a hybrid molecule with a first and a second recombined DNA sequences due to recombination between the first and the second DNA sequences to be recombined. Also another set of prokaryotic cells is obtained containing a hybrid molecule with a third and a fourth recombined DNA sequences due to recombination between the third and the fourth DNA sequences to be recombined. Then the first, second, third and fourth recombined DNA sequences thus obtained are isolated from their respective host cells. In the next step the first or the second recombined sequence can be inserted into a donor DNA molecule and the third or fourth recombined sequence into a recipient DNA molecule. Both the donor and recipient DNA molecules thus obtained are then introduced into a prokaryotic host cell and subjected another round of recombination. In this way, five, six or more diverging DNA sequences can also be recombined.

In a preferred embodiment of the invention the donor DNA molecule and the recipient DNA molecules are different linear or circular DNA structures. The term "DNA structure" means a DNA molecule, for example a vector such as a plasmid or bacteriophage, which is characterized in that it is present upon introduction into the prokaryotic host cell in the form of an extrachromosomal element. In the context of the present invention an "extrachromosomal element" is therefore a DNA molecule which does not integrate into the chromosome(s) of the prokaryotic host cell.

The donor DNA molecule must be able to hybridize with the recipient DNA molecule so that a hybrid molecule can be formed. Preferably the donor DNA molecule and the recipient DNA molecule do not share homology in general with the exception of the DNA sequences

5 to be recombined.

- According to the invention the recipient DNA molecule must be able to autonomously replicate in a prokaryotic host cell upon introduction into that cell. Therefore the recipient DNA molecule must have at least one origin of replication enabling the recipient DNA molecule to
- 10 replicate independently of the host genetic material. In the context of the present invention an "origin of replication" or "ori" is the region of a DNA molecule that is used by the cellular enzymes to initiate replication of that DNA molecule. At the origin, the two strands of DNA are pulled apart to form a replication bubble, which creates a region
- 15 of single-stranded DNA on each side of the bubble. The DNA polymerase machinery can then move in and begin to synthesize the new DNA strands, using the old strands as template. Small DNAs including bacterial plasmids or bacteriophages usually have a single origin.
- 20 In an embodiment of the invention the recipient DNA molecule is a plasmid, in particular a double-stranded circular DNA molecule. A "plasmid" is an extrachromosomal element which can replicate independently of the host genetic material. In another embodiment of the invention the recipient DNA molecule is a bacteriophage, in particular a phage the DNA of which is present in the prokaryotic cell in the form of an extrachromosomal element.
- 25

In a preferred embodiment of the invention the recipient DNA molecule is a plasmid, which can replicate in an *E. coli* host cell. Preferably, the recipient DNA molecule is the *E. coli* plasmid pACYC184 or a derivative thereof. Plasmid pACYC184 is Tet^R and Cam^R.

- 5 According to the invention the donor DNA molecule is a DNA molecule which cannot replicate in the prokaryotic host cell but is able to form a hybrid molecule with the recipient DNA. Therefore, any DNA molecule can be used as donor molecule as long as it contains at least a marker sequence and cannot independently replicate within a
- 10 given prokaryotic host cell. Examples of suitable donor molecule comprise, but are not limited to, linear double-stranded DNAs which can circularize, generated for example by PCR and containing an appropriate marker sequence, plasmids and bacteriophages. In case that a plasmid or a bacteriophage is used as donor DNA molecule
- 15 this molecule either does not have (a) functional origin(s) of replication at all or has (an) origin(s) of replication which is/are in the prokaryotic host cells used not functional, i.e. not active.

- 20 In one embodiment of the invention therefore the donor DNA molecule does not contain an origin of replication. This means, that the donor DNA molecule does not contain any sequences which could function in any prokaryotic host cell as an origin of replication, i.e. sequences to which protein factors involved in the initiation of replication could bind. The absence of an origin of replication can be due to a deletion of those nucleic acid sequences functioning as origin.
- 25 In another embodiment of the invention the function of the replication origin of the donor DNA molecule is impaired by one or more mutations abolishing either the function of the origin of replication itself or

the function of proteins involved in the replication, in particular those proteins that bind to the nucleic acid sequence of the origin whereby the replication is initiated. For example it is known, that in *E. coli* the key protein for initiation of replication, DnaA, binds to specific sequences, the so-called DnaA-boxes, at the chromosomal origin of replication and melts three AT-rich, 13-mer direct repeats. Additional binding of the DnaA protein to single-stranded 6-mer boxes in the open region is thought to stabilize the open complex (for a review see Jiang et al., PNAS, 100 (2003), 8692-8697). DnaA boxes and 5 AT-rich regions are commonly found at the origin of a variety of prokaryotic replicons and the DnaA protein has been shown to play a key role in the initiation of replication at these origins. Therefore, certain mutations of DnaA boxes and/or AT-rich regions or corresponding sites in the sequence of the replication origin, such as appropriate base substitutions, deletions etc., can prevent the binding of proteins required for initiation of replication and thus inhibit the function 10 of the origin of an extrachromosomal element. In a preferred embodiment of the invention therefore the function of the replication origin of the donor DNA molecule can be impaired by one or more mutations in the nucleic acid sequence of the replication origin of the donor DNA molecule, in particular in the DnaA boxes and/or AT-rich 15 regions of the origin.

20

Furthermore, it is known that the DnaA protein alone is not sufficient for the formation of an open complex at the origin of plasmids such 25 as RK2, P1, F, pSC101 or R6K. In these cases, efficient formation of an open complex requires binding of the plasmid Rep protein, although in concert with the host DnaA protein and HU or IHF (integrated host factor) protein. The requirement for a plasmid-specified initiation protein for open complex formation is key to the ability of

the plasmid to exercise control over the frequency of initiation events at its replication origin (for a review see Jiang et al., PNAS, 100 (2003), 8692-8697). This means, that mutations of plasmid nucleic acid sequences encoding protein factors with essential functions in

5 the plasmid replication can also impair the function of the replication origin of an extrachromosomal element. In another preferred embodiment of the invention therefore the function of the replication origin of the donor DNA molecule can be impaired by one or more mutations in nucleic acid sequences encoding protein factors with

10 essential functions in the replication of the donor DNA molecule.

In still another preferred embodiment of the invention the donor DNA molecule contains an origin of replication which is only active in certain prokaryotic host cells, but not in that prokaryotic cell into which the donor DNA molecule is introduced in order to effect recombination between the two DNA sequences to be recombined. There are numerous reports that an origin of replication that is active in a particular bacterial species does not function in another bacterial species. For example it is known that the replication origin of *Klebsiella pneumoniae* (oriC) is not active in *Caulobacter crescentus*, *Pseudomonas putida* or *Rhodobacter sphaeroides* (O'Neill and Bender, J. Bacteriol., 170 (1988), 3774-3777). It is also known that plasmids of *Bacillus subtilis* cannot replicate in *E. coli* cells. In a preferred embodiment therefore the donor DNA molecule and/or its origin of replication are derived from a prokaryotic species other than the prokaryotic species into the cells of which the donor DNA molecule is introduced.

In a particularly preferred embodiment the donor DNA molecule is the *B. subtilis* plasmid pMIX91, which is a derivative of plasmid

pIL253 and which can replicate in *B. subtilis*, but not in *E. coli*, and which comprises the *spec^R* marker, the *phleo^R* marker and the restriction sites *Scal*, *PpuMI* and *EcoO109I* for inserting a foreign DNA sequence. In another particularly preferred embodiment the donor 5 DNA molecule is the *B. subtilis* plasmid pMIX101, which is a derivative of plasmid pIL253 and which can replicate in *B. subtilis*, but not in *E. coli*, and which comprises the *tc^R* marker sequence and the restriction sites *XhoI* and *PstI* for inserting a foreign DNA sequence.

In another embodiment of the invention the replication origin of the 10 donor DNA molecule is functional only in a certain temperature range, for example at a temperature of less than 45°C. If the donor DNA molecule contains such a temperature sensitive origin it will not be able to replicate under non-permissive conditions, i.e. at a temperature above 45°C, which yet permit growth of the prokaryotic host 15 cell..

In the context of the present invention the term "marker sequences" refers to DNA sequences that are unique in a given prokaryotic cell and that are positioned on either the donor DNA molecule or the recipient molecule, preferably upstream or downstream of a recombination substrate or an already recombined DNA sequence. The presence of a marker sequence on the same molecule of DNA as the recombination substrate or already recombined DNA sequence, preferably in combination with another marker sequence, which can be positioned on the other side of the recombination substrate, allows that recombination substrate or already recombined DNA sequence to be recognized and selected for, whether by molecular or 20 genetic methods. 25

According to the invention the donor DNA molecule comprises a first marker sequence which allows for the selection of crossovers involving recombination substrates. The first marker sequence, preferably in combination with additional marker sequences present on either

- 5 the donor DNA molecule or the recipient DNA molecule, also allows further rounds of recombination to be carried out in a iterative fashion.

According to the invention the "first marker sequence" is a protein-encoding DNA sequence, the gene product of which is essential for

- 10 the prokaryotic host cell to grow and propagate under the selective conditions applied. The first marker sequence is selected from the group consisting of a nutritional marker, an antibiotic resistance marker and a sequence encoding a subunit of an enzyme which functions only, if both or more subunits are expressed in the same
15 cell.

A "nutritional marker" is a marker sequence that encodes a gene product that can compensate an auxotrophy of an organism or cell and thus can confer prototrophy on that auxotrophic organism or cell.

- In the context of the present invention the term "auxotrophy" means
20 that an organism or cell must be grown in a medium containing an essential nutrient which cannot be synthesized by the auxotrophic organism itself. The gene product of the nutritional marker gene promotes the synthesis of this essential nutrient missing in the auxotrophic cell. Therefore, upon expression of the nutritional marker
25 gene it is not necessary to add this essential nutrient to the medium in which the organism or cell is grown, since the organism or cell has acquired prototrophy.

An "antibiotic resistance marker" is a marker gene wherein the gene product confers upon expression to a cell, in which the expression of the antibiotic marker gene takes place, the ability to grow in the presence of a given antibiotic at a given concentration, whereas a

5 cell without the antibiotic resistance marker cannot.

A "sequence encoding a subunit of an enzyme" can be used as a marker sequence, if a cell cannot synthesize all subunits of an enzyme that are required for the assembly of the complete enzyme structure and thus for obtaining the full activity of the enzyme, and if

10 the presence or absence of the enzymatic activity can be monitored by genetic and/or molecular means. If, for example, the activity of an enzyme is needed for an essential biochemical pathway of the cell, which enables the growth and/or propagation of the cell in a particular environment, and the cell cannot synthesize all components of

15 the complete enzyme structure, then the cell cannot survive in that environment. The "sequence encoding a subunit of an enzyme" used as marker sequence therefore allows upon expression the assembly of the complete enzyme and the survival of the cell.

Preferably, the gene product of the first marker sequence confers

20 resistance to an antibiotic to a cell which is sensitive to that antibiotic. In particular, the first marker sequence is *spec^R* the gene product of which confers to a cell resistance to spectinomycin, *phleo^R* the gene product of which confers to a cell resistance to phleomycin or *tc^R* the gene product of which confers to a cell resistance to tetracycline.

25 In another embodiment of the invention the donor DNA molecule contains a second marker sequence. In still another embodiment of the invention the recipient DNA molecule contains a third marker se-

quence and optionally a fourth marker sequence. That means, in an embodiment of the invention the donor DNA molecule can comprise at least a first and a second marker sequence, optionally even more marker sequences, which can be arranged for example either upstream or downstream of the recombination substrate. In another embodiment the recipient DNA sequence can comprise a third and a fourth marker sequence, optionally even more marker sequences, which can be arranged for example either upstream or downstream of the recombination substrate. These additional marker allow to increase the stringency of selection of the recombined DNA sequences, since the hybrid molecule formed in the second prokaryotic cell and comprising the two different recombined DNA sequences can contain at least four different marker sequences altogether.

According to the invention the "second, third and fourth marker sequences" are protein-coding or non-coding sequences selected from the group consisting of nutritional markers, pigment markers, antibiotic resistance markers, antibiotic sensitivity markers, primer recognition sites, intron/exon boundaries, sequences encoding a particular subunit of an enzyme, promoter sequences, downstream regulated gene sequences and restriction enzyme sites.

A "pigment marker" is a marker gene wherein the gene product is involved in the synthesis of a pigment which upon expression will stain that cell, in which the pigment marker is expressed. A cell without the pigment marker does not synthesize the pigment and is therefore not stained. The pigment marker therefore allows a rapid phenotypical detection of that cell containing the pigment marker.

An "antibiotic sensitivity marker" is a marker gene wherein the gene product destroys upon expression the ability of a cell to grow in the presence of a given antibiotic at a given concentration.

- 5 "Primer recognition sites" are annealing sites for site-specific PCR primers, which allow a rapid identification of the respective marker sequences by PCR.

- In a preferred embodiment of the invention the second marker sequence of the donor DNA molecule and the third and fourth marker sequences of the recipient DNA molecule are protein-coding sequences, the gene products of which confer resistance to an antibiotic to a cell which is sensitive to that antibiotic. Preferably, the gene product of the third marker sequence confers to a cell resistance to tetracycline and the gene product of the fourth marker sequence confers to a cell resistance to chloramphenicol
- 10
- 15 In the context of the present invention the terms "DNA sequences to be recombined" and "recombination substrate" mean any two DNA sequences that can be recombined as a result of homologous or non-homologous recombination processes.

- Homologous recombination events of several types are characterized by the base pairing of a damaged DNA strand with a homologous partner, where the extent of interaction can involve hundreds of nearly perfectly matched base pairs. In contrast, illegitimate or non-homologous recombination is characterized by the joining of ends of DNA that share no or only a few complementary base pairs. In prokaryotic cells, non-homologous repair and recombination events occur at significantly lower frequencies than homologous recombination events.
- 20
- 25

In a preferred embodiment of the invention the two recombination substrates, i.e. the first and second DNA sequences to be recombined, are diverging sequences, i.e. sequences, which are not identical but show a certain degree of homology. In the context of the invention the term "homology" denotes the degree of identity existing between the sequences of two nucleic acid molecules, whereas "divergence" denotes the degree of non-identity between the sequences of two nucleic acid molecules. According to the invention the DNA sequences to be recombined differ from each other at two or more positions with respect to their overall alignment. In an embodiment of the invention the overall divergence between the two recombination substrates, relative to their total length, is more than 0,1 %, in particular more than 5 % and preferably more than 25 %. This means, that the DNA sequences to be recombined can differ by more than 30 %, by more than 40 % and even by more than 50 %. In a particularly preferred embodiment of the invention the two DNA sequences to be recombined share at least one or more homologous or identical regions, which, however, can be very short. According to the invention the homologous or identical regions can comprise less than 25 nucleotides, in particular less than 20 or less than 15 nucleotides and even less than 10 nucleotides, for example 4, 5, 6, 7, 8 or 9 nucleotides.

Recombination substrates or DNA sequences to be recombined can have a natural or synthetic origin. Therefore, in one embodiment of the invention the first and the second DNA sequences to be recombined are naturally occurring sequences. Naturally occurring sequences can be isolated from any natural source including viruses, living or dead prokaryotic organisms such as bacteria, living or dead eukaryotic organisms such as fungi, animals, plants and humans, or

parts thereof by any appropriate isolation methods or can be synthesized by chemical means. Naturally occurring sequences can also comprise such sequences that after isolation from a natural source were subjected to mutagenesis. In another embodiment of the invention the first and the second DNA sequences to be recombined are artificial sequences that are not found in a natural source. Artificial sequences can be synthesized by any known chemical methods.

In a preferred embodiment of the invention DNA sequences to be recombined are protein-encoding sequences, for example sequences encoding enzymes that can be utilized for the industrial production of natural and non-natural compounds. Enzymes or those compounds produced by the help of enzymes can be used for the production of drugs, cosmetics, foodstuffs, etc. Protein-encoding sequences can also be sequences that encode proteins that have therapeutic applications in the fields of human and animal health. Important classes of medically important proteins include cytokines and growth factors. The recombination of protein coding sequences allows for the generation of new mutated sequences which code for proteins with altered, preferably improved functions and/or newly acquired functions. In this way it is possible, for example, to achieve improvements in the thermostability of a protein, to change the substrate specificity of a protein, to improve its activity, to evolve new catalytic sites and/or to fuse domains from two different enzymes. Protein-encoding DNA sequences to be recombined can include sequences from different species which code for the same or similar proteins that have in their natural context similar or identical functions. Protein-encoding DNA sequences to be recombined can include sequences from the same protein or enzyme family. Protein coding sequences to be recombined can also be sequences which

code for proteins with different functions - for example, sequences that code for enzymes which catalyse different steps of a given metabolic pathway. In a preferred embodiment of the invention the first and the second DNA sequences to be recombined are selected

5 from the group of gene sequences of the Oxa superfamily of Beta-lactamases.

In another preferred embodiment of the invention DNA sequences to be recombined are non-coding sequences such as sequences, which, for example, are involved within their natural cellular context

10 in the regulation of the expression of a protein-coding sequence. Examples for non-coding sequences include, but are not limited to, promoter sequences, sequences containing ribosome binding sites, intron sequences, polyadenylation sequences etc. By recombining such non-coding sequences it is possible to evolve mutated sequences, which in a cellular environment result in an altered regulation of a cellular process - for example, an altered expression of a gene.

According to the invention a recombination substrate or DNA sequence to be recombined can of course comprise more than one protein coding sequence and/or more than one non-coding sequence. For example a recombination substrate can comprise one protein-encoding sequence plus one non-coding sequence or a combination of different protein-encoding sequences and different non-coding sequences. In another embodiment of the invention DNA sequences to be recombined therefore can consist of one or more stretches of coding sequences with intervening and/or flanking non-coding sequences. That means, the DNA sequence to be recombined can be for example a gene sequence with regulatory se-

- quences at its 5'-terminus and/or an untranslated 3'-region or an mammalian gene sequence with an exon/intron structure. In still another embodiment of the invention DNA sequences to be recombinant can consist of larger continuous DNA stretches that contain
- 5 more than a single coding sequence, optional with intervening non-coding sequences, such as an operon. DNA sequences to be recombinant can be sequences that have already experienced one or more recombination events, for example homologous and/or non-homologous recombination events.
- 10 The recombination substrates can comprise non-mutated wild-type DNA sequences and/or mutated DNA sequences. In a preferred embodiment therefore it is possible to recombine wild-type sequences with already existing mutated sequences in order to evolve new mutated sequences.
- 15 According to the invention a prokaryotic cell is used as host cell for introducing the recipient and donor DNA molecules. The terms "prokaryotic cell" and "prokaryotic host cell" include any cell, in which the genome is freely present within the cytoplasm as a circular structure, i.e. a cell, in which the genome is not surrounded by a nuclear mem-
- 20 brane. A prokaryotic cell is further characterized in that it is not necessarily dependant on oxygen and its ribosomes are smaller than that of eukaryotic cells. Prokaryotic cells include archaebacteria and eubacteria. In dependence on the composition of the cell wall eubacteria can be divided into gram-positive bacteria, gram-negative bacteria and cyanobacteria.
- 25

Therefore, according to the present invention the prokaryotic cell is a cell of an archaebacterium or an eubacterium, whereby in a pre-

ferred embodiment of the invention the prokaryotic cell is a gram-negative bacterium, a gram-positive bacterium or an cyanobacterium. Preferably, the gram-negative bacterium is *Escherichia coli*, for example the *E. coli* strain AB1157 or its MutS⁻ mutant, the *E. coli* strain MXP1.

According to the invention it may be preferred to use prokaryotic host cells for the inventive process which have a functional repair system. The mismatch repair (MMR) system is one of the largest contributors to avoidance of mutations due to DNA polymerase errors in replication. However, mismatch repair also promotes genetic stability by ensuring the fidelity of genetic recombination. Whereas in bacteria and also in yeast and mammalian cells, recombination between homologous DNA substrates containing a few mismatches (< 1%) occurs much less efficiently than between identical sequences, the frequency of recombination (gene conversion and/or crossovers) is dramatically elevated in MMR-defective lines. This means, that the high fidelity of recombination is not only caused by the intrinsic properties of recombination enzymes, but also by the editing of recombination by the mismatch repair system. Thus the mismatch repair machinery has an inhibitory effect on recombination between diverged sequence. In *E. coli* two proteins of the methyl-directed MMR system, namely MutS and MutL, are required for this strong antirecombination activity, whereas the effect of the other MMR system proteins, MutH and UvrD, is less pronounced. In addition to their roles in MMR and homeologous recombination, MMR proteins also play an important role in removing non-homologous DNA during gene conversion.

- In another preferred embodiment of the invention, prokaryotic cells that are deficient in the mismatch repair system are used. In the context of the present invention the term "deficient in the mismatch repair system" means that the MMR system of a prokaryotic cell is
- 5 transiently or permanently impaired. MMR deficiency of a cell or an organism can be achieved by any strategy that transiently or permanently impairs the MMR system including but not limited to a mutation of one or more genes involved in MMR, treatment with an agent like UV light, which results in a global impairment of MMR, treatment with an agent like 2-aminopurine or a heteroduplex containing an excessive amount of mismatches to transiently saturate and inactivate the MMR system, and inducible expression or repression of one or more genes involved in MMR, for example via regulatable promoters, which would allow for transient inactivation
- 10
- 15 In a preferred embodiment of the invention the mismatch repair deficiency of the prokaryotic host cell is due to a mutation of at least one of the genes involved in MMR. In a preferred embodiment the prokaryotic cells have a mutated *mutS* gene, a mutated *mutL* gene, a mutated *mutH* gene and/or a mutated *UvrD* gene.
- 20 In another embodiment the prokaryotic host cell is impaired or hampered in one or more of the major recombination proteins. It has been shown that a cell impaired for example in the *AddAB* genes shows a higher frequency of homologous and non-homologous recombinationIn another embodiment the prokaryotic host cell is over-expressing one of the major recombination proteins such as *recA*. This protein is involved in homologous recombination by promoting renaturation of single-stranded DNA to form the heteroduplex mole-
- 25

cule required for the recombination event and initiates the exchange of DNA strands.

According to the invention the first prokaryotic cell is generated by simultaneously or sequentially introducing the recipient DNA molecule and the donor recipient molecule into the prokaryotic host cell.

5 In one embodiment of the invention, therefore, the first prokaryotic cell can be generated by introducing in a first step the recipient DNA molecule into a particular prokaryotic host cell. After recovery of the prokaryotic host cell harbouring the recipient molecule, the donor

10 DNA molecule is introduced in a second step into the cell harbouring the recipient DNA molecule. In another embodiment of the invention both recipient and donor DNA molecules can be introduced simultaneously into the prokaryotic host cell.

According to the invention the introduction of both the recipient and

15 the donor DNA molecules into the prokaryotic host cell can be effected by any known appropriate method including, but not restricted to, transformation, conjugation, transduction, sexduction, infection and/or electroporation.

In the context of the present invention the term "transformation"

20 means the uptake of an isolated, preferably purified, nucleic acid molecule from the environment by a cell, for example a microbial cell. Cells of some prokaryotic species such as *Bacillus* or *Diplococcus* are naturally competent, whereas cells of other prokaryotic species such as *E. coli* have to be subjected special treatments in order

25 to make them competent, i.e. to induce the transfer of the nucleic acid molecule across the cellular membrane. Several bacteria are known for their ability to exchange DNA through transformation.

- "Conjugation" means the plasmid-mediated transfer of a bacterial plasmid from one bacterial cell into another bacterial cell through cell-to-cell-contact. The transfer machinery involved is usually encoded by plasmids or conjugative transposons. Examples of such
- 5 plasmids are conjugative plasmids or helper plasmids. Conjugative plasmids are self-transmissible plasmids that carry genes that promote cell-to-cell contact (mobilisation genes). They contain the genes for generating conjugation bridges. Once a bridge is made, other plasmids and even chromosomal DNA (conjugative transposons) can also be transferred. Mobilisable plasmids contain mobilisation genes, but need the "help" of conjugative plasmids to move between cells. Conjugation is one of the major routes for genetic exchange between different phylogenetic groups of prokaryote cells and between prokaryotes and eukaryotes.
- 10
- 15 "Sexduction" is a process by which genetic material is transferred from a prokaryotic cell with an F-factor or F-plasmid to a cell which does not contain that F-factor (F^- -cell). The F-factor is usually present in the cytoplasm of a bacterial cell, but can occasionally be incorporated at various sites of the bacterial chromosome, which leads
- 20 to the formation of Hfr-cells. The integration of the F-factor is reversible, whereby upon an incorrect disintegration of the plasmid so-called substituted F-factors (F') arise that can contain genetic material flanking in the bacterial chromosome the original integration site of the F-factor. The F' -plasmid can be transferred with high frequency into F^- -cells.
- 25

"Transduction" means the transfer of a nucleic acid molecule from one bacterial cell into another bacterial cell by a bacteriophage, which comprises the release of a bacteriophage from one cell and

the subsequent infection of the other cell. There are two types of transduction. A specialized transduction may occur during the lysogenic life cycle of a temperate bacteriophage, whereby genetic material of bacteria can replace a part of the bacteriophage genome. This 5 piece of bacterial DNA replicates as a part of the phage genome and can be transferred by the phage into another recipient cell. In case of a generalized transduction the entire genome of a lytic phage can be replaced by bacterial DNA. When this phage infects another bacterium, it injects the DNA into the recipient where it can be exchanged 10 for a piece of the DNA of the recipient cell.

"Electroporation" is a process where cells are mixed with nucleic acid molecules and then briefly exposed to pulses of high electrical voltage. The cell membrane of the host cell is rendered penetrable thereby allowing foreign nucleic acids to enter the host cell.

15 In a particular preferred embodiment of the invention the donor DNA molecule is UV-irradiated prior to introduction into the prokaryotic host cell, since it is known that irradiation leads to an increase of recombination frequencies.

According to the invention the first prokaryotic cell containing the recipient and donor DNA molecules is cultivated under such conditions 20 which force both the formation of a co-integrate or hybrid molecule between the recipient and donor DNA molecules and the recombination, preferably non-homologous recombination, between the two recombination substrates, i.e. conditions which allow for the selection 25 of recombinant DNA sequences.

If the first marker sequence of the donor DNA molecule is an antibiotic resistance marker, the first prokaryotic cell containing the recipi-

- ent and donor DNA molecules is cultivated in the presence of an antibiotic to which the prokaryotic host cell is sensitive and to which the gene product of the first marker sequence confers resistance. In a particularly preferred embodiment the first prokaryotic is cultivated in
- 5 the presence of spectinomycin if the first marker sequence present on the donor DNA molecule is *spec^R* the gene product of which confers to a cell resistance to spectinomycin. In another particularly preferred embodiment the first prokaryotic is cultivated in the presence of phleomycin if the first marker sequence present on the donor DNA
- 10 molecule is *phleo^R* the gene product of which confers to a cell resistance to phleomycin. In another particularly preferred embodiment the first prokaryotic is cultivated in the presence of tetracycline if the first marker sequence present on the donor DNA molecule is *tc^R* the gene product of which confers to a cell resistance to tetracycline.
- 15 If the first marker sequence of the donor DNA molecule is a nutritional marker, the first prokaryotic cell containing the recipient and donor DNA molecules is cultivated in a medium lacking a particular essential nutrient which cannot be synthesized by the host cell itself and which is supplied to the host cell upon introduction of the donor
- 20 molecule and expression of the first marker gene contained on the donor molecule. Upon expression of the first nutritional marker gene it is not necessary to add this essential nutrient to the medium in which the prokaryotic cell is grown.

If the second marker sequence, the third marker sequence and/or the fourth marker sequence are antibiotic resistance markers then in a preferred embodiment the first prokaryotic cell can be cultivated additionally in the presence of a second, a third and/or a fourth antibiotic to which the gene products of the second marker sequence,

the third marker and the fourth marker sequence, respectively, confer resistance. Preferably, the first prokaryotic cell is cultivated not only in the presence of phleomycin or spectinomycin or tetracycline, but in addition also in the presence of chloramphenicol. Most preferably, the first prokaryotic cell is grown and propagated in presence of phleomycin, spectinomycin, chloramphenicol and tetracycline.

- After cultivation of the prokaryotic cells under selective conditions second prokaryotic cells are isolated which contain a co-integrate or hybrid molecule formed by the recipient and donor DNA molecules.
- 10 This co-integrate contains the recombinant DNA sequences. The presence of the co-integrate and/or recombinant DNA sequences can be verified and detected by several means such as restriction profile analysis, PCR amplification and/or sequencing. If, for example the second marker sequence of the donor molecule and the third or
- 15 fourth marker sequence of the recipient molecule are unique primer recognition sequences then specific fragments of the co-integrate can be PCR-amplified by using appropriate primers recognizing these marker sequences. In order to detect the presence of the respective marker combination. If no co-integrate has formed, i.e. no
- 20 recombination has occurred, these fragments cannot be detected. If, for example the second marker sequence of the donor molecule and the third or fourth marker sequence of the recipient molecule are unique restriction enzyme cleavage sites then the co-integrate can be subjected to a restriction enzyme analysis in order to detect specific DNA fragments. If no co-integrate has formed, i.e. no recombination has occurred, these fragments cannot be detected.
- 25

According to the invention the first and the second recombined DNA sequences contained in the hybrid DNA molecule of the second pro-

karyotic cell can be isolated and/or analysed and/or selected. The first and second recombined DNA sequences obtained can be isolated for example by PCR amplification or cleavage with appropriate restriction enzymes. An analysis of the first and the second recom-

- 5 combined DNA sequences contained in the hybrid DNA molecule can be conducted for example by sequencing methods.

According to the invention the isolated first and second recombined DNA sequences can again be inserted into a donor DNA molecule and a recipient DNA molecule, respectively, and subjected to another round of recombination by using the inventive process.

- 10 Another aspect of the present invention relates to a process of generating novel proteins, enzymes and non-coding sequences with novel or improved functions and properties, whereby known protein-coding sequences or known non-coding sequences are subjected
15 15 one or more recombination rounds by using the inventive process for generating and detecting recombinant DNA sequences in a prokaryotic host cell. The present invention also relates to the proteins, enzymes and non-coding sequences, generated by any one of the inventive processes.

- 20 The present invention also relates to the *B. subtilis* plasmid pMIX91, which is a derivative of plasmid pIL253 and which can replicate in *B. subtilis*, but not in *E. coli*, and which comprises the *spec^R* marker and the *phleo^R* marker and the restriction sites *Scal*, *PpuMI* and *EcoO109I* for inserting a foreign DNA sequence.

- 25 The present invention also relates to the *B. subtilis* plasmid p-MIX101, which is a derivative of plasmid pIL253 and which can replicate in *B. subtilis*, but not in *E. coli*, and which comprises the *tc^R*

marker sequence and the restriction sites *Xhol* and *PstI* for inserting a foreign DNA sequence.

The present invention also relates to *B. subtilis* strain DSM4393, containing the *B. subtilis* plasmid pMIX91 and to *B. subtilis* strain 5 1A423, containing the *B. subtilis* plasmid pMIX101.

Another aspect of the present invention relates to the use of the *B. subtilis* plasmid pMIX91 and pMIX101, respectively, as donor DNA molecules in the inventive process, i.e. the use of plasmid pMIX91 or pMIX101 for generating and/or detecting recombinant DNA sequences in a prokaryotic host cell, preferably in an *E. coli* cell. 10

Another aspect of the present invention relates to the use of the *E. coli* plasmid pACYC184 or pMIX100 or derivatives thereof as recipient DNA molecule in the inventive process, i.e. the use of plasmid pACYC184 or pMIX100 for generating and/or detecting recombinant 15 DNA sequences in a prokaryotic host cell, preferably in an *E. coli* cell.

Another aspect of the present invention relates to a kit which can be used for conducting the inventive process for generating and detecting recombined DNA sequences in a prokaryotic host cell. In a first 20 embodiment the kit comprises at least a first container which comprises cells of the *E. coli* strain AB1157 as prokaryotic host cell, a second container which comprises cells of the *E. coli* strain AB1157, containing the *E. coli* plasmid pACYC184 or the *E. coli* plasmid pMIX100, which can be used as recipient DNA molecule, and a third 25 container comprising cells of the *B. subtilis* strain DSM4393, containing the *B. subtilis* plasmid pMIX91, or cells of the *B. subtilis* strain 1A423, containing the *B. subtilis* plasmid pMIX101, which can be

used as donor DNA molecule. In a second embodiment of the invention the kit comprises at least a first container which comprises cells of the *E. coli* strain MXP1, which is a MutS⁻ mutant of strain AB1157, as prokaryotic host cell, a second container which comprises cells of
5 the *E. coli* strain AB1157 containing plasmid pACYC184 or pMIX100 and a third container comprising cells of the *B. subtilis* strain DSM4393, containing the plasmid pMIX91, or cells of the *B. subtilis* strain 1A423, containing the plasmid pMIX101. In still another embodiment the kit comprises at least a first container which comprises
10 either cells of the *E. coli* strain AB1157 or the *E. coli* strain MXP1, a second container comprising DNA of the *E. coli* plasmid pACYC184 or pMIX100 and a third container comprising DNA of the *B. subtilis* plasmid pMIX91 or the *B. subtilis* plasmid pMIX101.

Another aspect of the present invention relates to a process for producing a hybrid gene and/or a protein encoded by a hybrid gene in a prokaryotic cell. The process for producing a hybrid gene and/or the encoded protein comprises the step of conducting the inventive process for generating and detecting recombinant DNA sequences, whereby the hybrid gene and/or the protein encoded by the hybrid
15 gene is produced in the prokaryotic cell. After expression the hybrid gene and/or the encoded protein is selected in the prokaryotic cell
20 and/or isolated therefrom.

The present invention also relates to a hybrid gene that is obtainable by the inventive process for producing a hybrid gene or the inventive process for generating and detecting recombinant DNA sequences.
25

The present invention also relates to a protein, which is encoded by a hybrid gene, that is obtainable by the inventive process for produc-

ing a hybrid gene or the inventive process for generating and detecting recombinant DNA sequences, and/or which is obtainable by the inventive process for producing a protein encoded by a hybrid gene.

5 The present invention is illustrated by the following sequence listing,
figures and example.

Figure 1 illustrates schematically the inventive process for generating and/or detecting recombination between two DNA sequences. The first sequence, the *oxa7* gene, is present on the *B. subtilis* plasmid pTG2-phleo as donor DNA molecule. pTG2-phleo carries the *spec^R* and *phleo^R* markers which confer resistance to spectinomycin and phleomycin. By electroporation pTG2-phleo is introduced into a *E. coli* host cell containing the plasmid pTG3 as recipient DNA molecule. pTG3 contains the second DNA sequence, the *oxa11* gene, and the *cm^R* marker conferring resistance to chloramphenicol. After 10 introduction of pTG2-phleo the cell is cultivated in presence of spectinomycin and phleomycin which forces the formation of a co-integrate between the two plasmids and concomitantly recombination between the genes. Therefore, upon cultivation an *E. coli* cell is obtained containing a dimeric plasmid containing the newly recombined 15 DNA sequences R1 and R2. The recombined DNA sequences can be analyses by restriction profile analysis, PCR amplification of R1 and R2 recombinant genes and/or sequencing of R1 and R2.
20
25

Figure 2 shows physical maps of the plasmids pUC19-phleo, pic156, pACYC184 and pIL253 used to construct suitable plasmids for conducting the inventive process.

Figure 3 shows physical maps of plasmids which were constructed for conducting the inventive process. The plasmids pMIX96 and

pMIX 97 are not depicted since they are as pMIX95 but with *oxa5* and *oxa1*, respectively, instead of *oxa11*. pMIX99 is as pMIX98, but with *oxa1* instead of *oxa11*.

Figure 4 shows the structure of genes obtained by the inventive in vivo recombination process between 22%-divergent *oxa* genes.
5 Stretches of sequence identity in which crossovers took place are detailed.

Figure 5 shows the structure of the *E. coli* plasmid pMIX100 which can be used as recipient DNA molecule in an *E. coli* host cell.
10 PMIX100 carries the origin of replication from plasmid pACYC184, as well as the chloramphenicol-resistance gene therefrom. PMIX100 also carries the gene *lacZ* from pBluescript SK+.

Figure 6 shows the structure of the *B. subtilis* plasmid pMIX101 which is a derivative of the *B. subtilis* plasmid pIL253 and which can be used as donor DNA molecule in an *E. coli* host cell. pMIX101 carries the marker *Erm*^R which confers resistance to erythromycin and *Tc*^R which confers resistance to tetracycline. The tetracycline-resistance gene was amplified from plasmid pACYC184.
15

Figure 7 shows the structure of the transformation efficiency control plasmid pMIX102. pMIX102, a derivative of pBluescript SK+, contains the *Tc*^R gene amplified from plasmid pACYC184. In pMIX102 the *Tc*^R gene is driven by the *plac* promotor
20

Figure 8 shows the structure of the transformation efficiency control plasmid pMIX103. pMIX103, a derivative of pBluescript SK+, contains the *Tc*^R gene amplified from plasmid pACYC184. In pMIX103
25

the Tc^R gene is cloned in the opposite direction than plac . Therefore, the gene is expressed from its own promoter.

Figure 9 illustrates schematically the strategy for cloning the $oxa7$, $oxa11$ and $oxa5$ genes into the *E. coli* plasmid pMIX100 and into the *B. subtilis* plasmid pMIX101. For cloning $oxa7$, $oxa11$ and $oxa5$ into pMIX100 the genes were amplified using primers containing either *PstI* or *Xhol* at their 5' ends. After digestion with those enzymes, the amplified DNA fragments were independently ligated with pMIX100, which was previously cut with *PstI* + *Xhol*. Competent cells of *E. coli* DHB10 were electroporated with the ligation mixtures, and colonies harboring positive clones were phenotypically selected by chloramphenicol-resistance/white color on LB plates containing Cm (30 µg/ml) + X-Gal (80 µg/ml) + IPTG (0.5 mM). Plasmid DNAs were obtained and analysed by restriction mapping. Results confirmed that pMIX104 carries $oxa7$, pMIX106 carries $oxa11$ and pMIX107 carries $oxa5$. For cloning into the *B. subtilis* plasmid pMIX101 $oxa7$ was obtained as a 0.9-kb fragment from pMIX104 by restriction with *PstI* and *Xhol* and ligated with pMIX101, which was previously cut with the same enzymes. After transformation of *B. subtilis* 1A423 competent cells with the ligation product, cells were selected in LB containing 0.5 µg/ml of erythromycin (Erm).

SEQ ID No. 1 and 2 show the sequences of the primers OLG1 and OLG2, respectively, for the amplification of $oxa7$ and the introduction of *Scal* and *PpuMI* restriction sites at the 5' and 3' ends, respectively. SEQ ID No. 3 and 4 show the sequences of the primers OLG3 and OLG4, respectively, for the amplification of $oxa11$ and the introduc-

tion of *Bam*HI and *Eco*O109I restriction sites at the 5' and 3' ends, respectively.

SEQ ID No. 5 and 6 show the sequences of the primers OLG5 and OLG6, respectively, for the amplification of *oxa*5 and the introduction of *Bam*HI and *Eco*O109I restriction sites at the 5' and 3' ends, respectively.

SEQ ID No. 7 and 8 show the sequences of the primers OLG7 and OLG8, respectively, for the amplification of *oxa*1 and the introduction of *Bam*HI and *Eco*O109I restriction sites at the 5' and 3' ends, respectively.

SEQ ID No. 9 and 10 show the sequences of the primers OLG9 and OLG10, respectively, for the amplification of *oxa*11 and the introduction of *Sc*aI and *Eco*O109I restriction sites at the 5' and 3' ends, respectively.

SEQ ID No. 11 shows the sequence of primer OLG11, which was used together with primer OLG8 (SEQ ID No.8) for the amplification of *oxa*1 and the introduction of *Sc*aI and *Eco*O109I restriction sites at the 5' and 3' ends, respectively.

SEQ ID No. 12 and 13 show the sequences of the primers OLG12 and OLG13, respectively, for the amplification of the recombinant gene R1 contained on the hybrid plasmid pMIX93.

SEQ ID No. 14 shows the sequence of primer OLG14, which was used together with primer OLG12 (SEQ ID No. 12) for the amplification of the recombinant gene R1 contained on one of the hybrid plasmids pMIX95, pMIX96 and pMIX97.

SEQ ID No. 15 shows the sequence of primer OLG15, which was used together with primer OLG17 (SEQ ID No. 17) for the amplification of the recombinant gene R2 contained on the hybrid plasmid pMIX93.

- 5 SEQ ID No. 16 shows the sequence of primer OLG16, which was used together which was used together with primer OLG17 (SEQ ID No. 17) for the amplification of the recombinant gene R2 contained on one of the hybrid plasmids pMIX95, pMIX96 and pMIX97.

SEQ ID No. 17 shows the sequence of primer OLG17.

10 **Example 1**

In vivo recombination of heterologous genes using a double plasmid system (E. coli/B. subtilis) whereby recombinant DNA sequences are selected by resistance to spectinomycin and/or phleomycin

15 1. **Material and methods**

1.1 Bacterial strains and plasmids

Bacterial strains and plasmids used in this example are shown in Table 1 and Table 2, respectively.

Table 1: Bacterial strains

| Strains | Genotype | Reference or source |
|--|--|--|
| <i>E. coli</i> AB1157 Nal^R hsd^- | $\text{thr1 leu6 proA2 his4,}$ thil argE3 lacY1 galK2 ara14 xyl15 mtl1 tsx33 str31 $\text{supE44thr}^+ \text{ hsdR}^- \text{B}$ nal^R | M. Radman strain collection |
| <i>E. coli</i> MIXP1 | As AB1157 $\text{Nal}^R R^-$ but $\text{mutS::Tn5 (kan}^R)$ | mutS allele from M. Radman strain collection |
| <i>E. coli</i> DH5 α | $\text{SupE44 } \Delta\text{lacU169 (8)}$ ($\phi 80\text{lacZ } \Delta\text{M15}$) $\text{hsdR17 recA1 endA1}$ $\text{gyrA96 thi-1 rel A1}$ | |
| <i>B. subtilis</i> DSM4393 | aroB2 trpC2 his6 | German Strains collection DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH) |

Table 2: Plasmids

| Plasmids | Selection | Reference or source |
|-------------|-------------------------------------|----------------------------|
| pACYC184 | Tet ^R Cam ^R | New England Biolabs (1) |
| pIL253 | Erm ^R | (6) |
| pic156 | Amp ^R Spec ^R | (7) |
| pUC19-Phleo | Amp ^R Phleo ^R | (2) |

1.2 Growth conditions and culture media

Both *E. coli* and *B. subtilis* strains were cultured at 37° C in LB medium (Difco Laboratories, Detroit, USA). When medium was used in plates (LBA), 15 g of agar (Difco) was added per liter. When required medium was supplemented with antibiotics the final concentrations were as follows: tetracycline (Sigma-Aldrich Chimie, St. Quentin Fallavier, France), 12.5 µg/ml; chloramphenicol (Sigma-Aldrich Chimie), 5 30 µg/ml; ampicillin (Sigma-Aldrich Chimie), 100 µg/ml; erythromycin (Sigma-Aldrich Chimie), 0.5 µg/ml; spectinomycin (Sigma-Aldrich Chimie), 75 µg/ml; phleomycin (Euromedex, Strasbourg, France), 10 2 µg/ml. When LBA was supplemented with spectinomycin plus phleomycin the final concentrations were 60 and 1 µg/ml, respectively.

15

1.3 Manipulation of DNA and microorganisms

Transduction

E. coli MIXP1 strain was constructed using P1-mediated transduction (3).

5 Transformation

Plasmids were introduced in *E. coli* strains by electrotransformation using an Eppendorf Electroporator 2510 (Eppendorf AG, Hamburg, Germany) and according to the supplier's instructions.

Competent cells of *B. subtilis* were prepared and transformed as described by Yasbin et al. (9)

DNA manipulations

Established protocols were followed for molecular biology techniques (4). Enzymes for DNA manipulations were purchased from New England Biolabs (Beverly, MA, USA), MBI Fermentas (Vilnius, Lithuania),
15 Promega (Madison, Wis.) or Stratagene (La Jolla, CA, USA) and used as recommended by the manufacturers. When necessary, DNA digested with restriction endonucleases was purified from agarose gels using the NuceloSpin Extract Kit (Machery-Nagel).

Plasmid DNA was isolated from *E. coli* using the NucleoSpin kit
20 (Machery-Nagel GmbH & Co., Düren, Germany) according to the instructions of the manufacturer. To isolate plasmid DNA from *B. subtilis* the same kit was used but a first lysis step was performed by incubating cells with 2 mg ml⁻¹ of lysozyme 30 min at 37°C.

Primers used were synthesized by Proligo France SAS (Paris, France).

Nucleotide sequences were determined in both directions by Genome Express (Meylan, France). Sequences were analysed with the 5 Infobiogen package (Genopole d'Evry, Evry, France). The ClustalW program was used for sequence comparisons.

PCR amplifications

PCR reactions were performed using a Mastercycler Gradient (Eppendorf AG, Hamburg, Germany). Reactions were carried out in 50-10 µl volume using the high-fidelity Herculase Enhanced DNA polymerase (Stratagene) under the following conditions: 96°C for 3 min, 35 cycles of 96°C for 30 s, annealing temperature for 30 s, 72°C for 1 min, and a final elongation step at 72°C for 10 min. Annealing temperature was determined by subtracting 5°C to the lowest Tm of primers used. When necessary, PCR products were purified using the 15 NucleonSpin Extract kit (Machery-Nagel). Amplification products were analysed by electrophoresis in 0.7% agarose gels (Sigma).

2. General strategy

This experiment was conducted in order to generate new molecules 20 exhibiting advantageous properties by *in vivo* recombination of two parental genes, which share different degrees of sequence identity. The strategy employed is as follows:

Genes that will be recombined are carried by two different plasmids showing no homology in their nucleotide sequences. The first one is 25 an *E. coli*-replicative plasmid conferring resistance to either chloram-

phenicol (Cm) or tetracycline (Tc). It is based on the standard cloning vector pACYC184, a low-copy number plasmid obtained from New England Biolabs.

- The second one is a *Bacillus subtilis* plasmid derived from pIL253
5 (Simon and Chopin, 1988). It is not able to replicate in *E. coli* and carries two antibiotic resistance markers for spectinomycin (Spc) and phleomycin (Phleo), respectively.

To recombine pairs of heterologous genes carried by these two vectors, the *B. subtilis* plasmid is introduced by electroporation into *E. coli* strains harbouring the replicative plasmid. This is shown schematically in Figure 1. Such strains are either proficient (+) or deficient (-) for the mismatch repair system (MMR), which controls both mutagenesis and recombination.
10

After electroporation, transformants are selected with antibiotics for
15 which the *B. subtilis* plasmid confers resistance (Spc and Phleo). Such selective pressure force recombination between heterologous genes, since only cells harbouring a co-integrate formed between the *B. subtilis* and the *E. coli* plasmids can grow under those conditions. Hybrid plasmid confers resistance to Spc and Phleo and replicates
20 from the *E. coli* origin, since that of *B. subtilis* is not functional. Moreover, it carries the two recombinant genes R1 and R2.

The first step was the cloning in the *E. coli* and *B. subtilis* vectors of genes initially chosen as target to evaluate recombination efficiency.

Recombination experiments between oxa genes were performed in
25 both wild type and mismatch repair deficient strains. Experiments were done between pairs of either identical or divergent genes.

Plasmid DNAs were UV-irradiated prior to electroporation, since it was shown that irradiation leads to a 10-fold increase of recombination frequencies.

Recombination was also performed in strains temporarily rendered 5 mutator by treatment with 2-aminopurine. 2-aminopurine is an adenine analog that is incorporated into DNA during growth of bacteria and saturates the MMR system. Thus, a transient mutator phenotype is generated, since after removing of 2-aminopurine a wild type status is recovered. This temporal control of the mismatch repair activity provides a stable background to strains used in recombination 10 avoiding the accumulation of mutations in their genomes.

3. Results

Construction of the *B. subtilis* vector

In order to construct a *B. subtilis* vector to carry target genes for recombination, two gene markers conferring resistance to antibiotics 15 spectinomycin (*spec*^R) and phleomycin (*phleo*^R), respectively, were cloned in the plasmid pIL253 following a two steps cloning strategy.

First, the *spec*^R gene was obtained as a SacI fragment of 1294 bps in length from the plasmid pic156. The fragment was purified and 20 then ligated to SacI-digested pIL253. *B. subtilis* DS4393 competent cells were transformed with the ligation mixture and transformants were selected on LBA plates containing 75 µg ml⁻¹ of spectinomycin. Restriction analyses of plasmid DNA obtained from transformants confirmed that they harboured pILF253-derivatives carrying *spec*^R 25 gene.

In a second step, the *phleo*^R gene was cloned into pIL253-spec. Plasmid pUC19-phleo was digested with *Eco*RI and *Sal*I restriction enzymes and the fragment of 574 bps in length corresponding to *phleo*^R was gel-purified and ligated to pIL253-spec, previously digested with the same two enzymes. *B. subtilis* DSM4393 competent cells were transformed with the ligation mixture and transformants were then selected on LBA plates containing 60 µg ml⁻¹ of phleomycin. Restriction analyses of plasmid DNA obtained from transformants demonstrated that they harboured the expected 6.69-kb plasmid, carrying both *spec*^R and *phleo*^R genes. Plasmid was named pMIX91 (see figure 3).

Construction of transformation efficiency control vector

A vector was constructed to be used as a control of transformation efficiency of *E. coli* strains under spectinomycin and phleomycin selection in recombination experiments. The vector was constructed as follows: the spectinomycin resistance gene (*spec*^R) was obtained as a fragment of 1.25 kb after digestion of pic156 with *Bam*H1 and *Eco*RI. It was cloned into the corresponding sites of pUC-phleo, next to the phleomycin resistance gene (*phleo*^R). After electroporation of *E. coli* DH10B competent cells with the ligation mixture, spectinomycin and phleomycin resistant colonies were selected on LBA plates containing both antibiotics at final concentrations of 60 µg/ml and 1 µg/ml, respectively. Restriction analyses of plasmid DNA isolated from transformants showed that they corresponded with the expected construct of 4.46 kb in length, which was named pMIX92 (see figure 3).

Cloning of genes encoding for β-lactamases into *E. coli*- and *B. subtilis* vectors

Four genes encoding for β-lactamases were chosen as target to evaluate recombination efficiencies in both wild-type and MutS⁺ mutant *E. coli* strains. Such genes, *oxa7* (GenBank accession number X75562), *oxa11* (GenBank accession number Z22590), *oxa5* (GenBank accession number X58272) and *oxa1* (GenBank accession number J02967) display different degrees of divergence in their nucleotide sequences. The sequence of *oxa1* and the sequences of 5 *oxa5*, *oxa7* and *oxa11*, respectively, diverge by 40%. The sequence of *oxa5* and the sequences of *oxa7* and *oxa11*, respectively, diverge by 22%. The sequence of *oxa7* and the sequence of *oxa11* diverge by 10 5%. 20

The four *oxa* genes were cloned into the *E. coli* plasmid pACYC184, whereas *oxa7*, *oxa11* and *oxa1* were cloned into the *B. subtilis* 15 plasmid pMIX91 as well. The cloning of the genes was done as follows: *oxa7* was amplified by PCR with primers designed to introduce *Sca*I and *PpuMI* sites at the 5' and 3' ends, respectively, of the amplified DNA (primers OLG1 with SEQ ID No. 1 and OLG2 with SEQ 20 ID No. 2). The PCR product was digested with those restriction enzymes and the resulting fragment of 991 bps in length was ligated to pMIX91, previously cut with the same enzymes. *B. subtilis* DSM4393 competent cells were transformed with the ligation mixture and transformants selection was done on LBA plates containing 75 µg ml⁻¹ of 25 spectinomycin. Restriction analyses of plasmid DNA obtained from transformants demonstrated that they harboured the expected 6.89 kb plasmid pMIX94 carrying *oxa7* (see figure 3).

To clone *oxa7* into the *E. coli* vector pACYC184, the PCR product described above, as well pACYC184, were digested with *PpuMI* and *Scal*. Blunt ends were generated from digested DNAs by using the Klenow fragment of DNA polymerase I, and ligation between them was done. *E. coli* DH10B competent cells were electroporated with the ligation mixture, and transformants were selected on LBA plates containing 12.5 µg ml⁻¹ of tetracycline. Restriction analyses of plasmid DNA isolated from transformants showed that they corresponded with the expected construct of 4.33 kb in length, which was named pMIX93 (see figure 3).

To clone *oxa11*, *oxa5* and *oxa1* into pACYC184, genes were amplified by PCR using primers designed to introduce *BamHI* and *EcoO109I* sites at the 5' and 3' ends, respectively, of the amplified DNA. OLG3 (SEQ ID No. 3)/OLG4 (SEQ ID No. 4), OLG5 (SEQ ID No. 5)/OLG6 (SEQ ID No. 6) and OLG7 (SEQ ID No. 7)/OLG8 (SEQ ID No. 8) primers pairs were used to amplify *oxa11*, *oxa5* and *oxa1* respectively. The PCR products were digested with *BamHI* and *EcoO109I* and the resulting fragment of 997 bps (*oxa11*) and 830 bps (*oxa5*) and 936 bps (*oxa1*) were independently ligated to pACYC184, previously cut with the same enzymes. *E. coli* DH10B competent cells were electroporated with the ligation mixtures, and transformants were selected on LBA plates containing 30 µg ml⁻¹ of chloramphenicol. Restriction analyses of plasmid DNA isolated from transformants showed that they corresponded the expected 6.89 kb plasmids pMIX95 (3.72 kb), pMIX96 (3.55 kb) and pMIX97 (3.66 kb) (see figure 3).

To clone *oxa11* and *oxa1* into pMIX91, genes were amplified by PCR using primers designed to introduce *Scal* and *EcoO109I* sites at the

5' and 3' ends, respectively, of the amplified DNA. OLG9 (SEQ ID No. 9)/OLG10 (SEQ ID No. 10), and OL11 (SEQ ID No.11)/OLG8 primers pairs were used to amplify *oxa11* and *oxa1* respectively. The PCR products were digested with *Sca*I and *EcoO109*I and the resulting fragment of 995 bps (*oxa11*) and 934 bps (*oxa1*) were independently ligated to pMIX91, previously cut with the same enzymes. *B. subtilis* DSM4393 competent cells were transformed with the ligation mixture and transformants selection was done on LBA plates containing 75 µg ml⁻¹ of spectinomycin. Restriction analyses of plasmid 10 DNA obtained from transformants demonstrated that they correspond to the expected plasmids pMIX98 (6.89kb) and pMIX99 (6,83 kb). See figure 3.

In vivo recombination of *oxa* genes in wt compared to MutS⁻ mutant *E.coli*

15 In a first step, competent cells of *E. coli* AB1157 and its MutS⁻ mutant, *E. coli* MXP1, were independently transformed by electroporation with pACYC184-derivative plasmids carrying *oxa* genes, pMIX93, pMIX95, pMIX96 or MIX97. Transformants were selected on the basis of their resistance to either tetracycline or chloramphenicol. The presence of the suitable plasmids was subsequently confirmed by restriction and/or PCR analyses.
20

25 In a second step, competent cells of both wild-type and MutS⁻ strains harbouring replicative plasmids were prepared from selective media containing either tetracycline or chloramphenicol. Afterwards such competent cells were independently transformed by electroporation with the *B. subtilis*-plasmids carrying *oxa* genes, pMIX94, pMIX98 or pMIX99.

After electroporation, transformants were selected on LBA plates containing antibiotics for which the *B. subtilis* plasmid confers resistance: spectinomycin and phleomycin, at final concentrations of 60 µg ml⁻¹ and 1 µg ml⁻¹, respectively. Such selective pressure force
5 recombination between oxa genes, since only cells harbouring a hybrid plasmid formed between the *B. subtilis* and the *E. coli* plasmids can grow in those conditions. Plates were incubated overnight at 37° C, and afterwards plasmid DNA of transformants were analyzed by digestion with restriction enzymes in order to confirm that they har-
10 boured hybrid plasmids (of about 10.5 kb in length) carrying two recombinant genes, R1 and R2. In some cases, recombinant genes were amplified by PCR and sequenced. If the resident plasmid was pMIX93 R1 was amplified with OLG12 (SEQ ID No. 12)/OLG13 (SEQ ID No. 13) and R2 with OLG15 (SEQ ID No. 15)/OLG17 (SEQ
15 ID No. 17); if the resident plasmid was pMIX95, pMIX96 or pMIX97 R1 was amplified with OLG12/OLG14 (SEQ ID No. 14) and R2 with OLG16 SEQ ID No. 16)/OLG17.

In recombination experiments the *B. subtilis* plasmid pMIX91, was used as a negative control, whereas the *E. coli* plasmid pMIX92 was
20 used as control of transformation efficiency under spectinomycin and phleomycin selection. pMIX92 can replicate in strains harbouring pACYC184-derivative plasmids since their origins of replication (ColE1 and p15, respectively) are compatible. Plasmid DNAs were UV-irradiated prior to electroporation (200J/m²), since it was shown
25 that irradiation leads to a 10-fold increase of recombination frequencies.

Recombination frequencies were calculated by dividing the transformation efficiency obtained with *B. subtilis* plasmids per the transfor-

mation efficiency obtained with control vector pMIX92. Transformation efficiencies were in turn calculated as number of colony forming units (cfu) obtained per μ g of DNA on previously described conditions.

- 5 Results obtained are summarized in table 3. In the wild-type strain, recombinants were obtained in experiments using either identical or 5% divergent oxa genes, whereas in the MutS⁻ mutant, recombination also happened between 22% divergent genes.
- 10 Table 3: In vivo recombination frequencies obtained between oxa genes in both wild-type and MutS⁻ *E. coli* strains.

| Genetic divergence (%) | Recombination frequencies | |
|---------------------------|---------------------------|--------------------------|
| | wt strain | MutS ⁻ strain |
| 0 | 10^{-4} | 10^{-4} |
| 5 | 10^{-8} | 10^{-5} |
| 22 | --- | 10^{-7} |
| 40 | --- | --- |

In vivo recombination of oxa genes in wt compared to 2-aminopurine (2-AP) treated *E.coli*

In a first step, competent cells of *E. coli* AB1157 were transformed by electroporation with pACYC184-derivative plasmids carrying *oxa11* gene, pMIX95. Transformants were selected on the basis of their resistance. The presence of the suitable plasmids was confirmed by restriction and/or PCR analyses.

5 In a second step, competent cells of this strain were prepared in presence of 200 µ g/ ml of 2-AP and independently electroporated:
10 with the *B. subtilis* plasmids pMIX94 carrying *oxa7* or pMIX98 carrying *oxa11*.

The *B. subtilis* plasmid pMIX91 was used as a negative control, whereas the *E. coli* plasmid pMIX92, carrying Spc^R and Phleo^R markers, was used as control of transformation efficiency. Plasmid DNAs
15 were UV-irradiated before electroporation in order to increase recombination frequency.

Results are summarized in the table 4. In the wild-type strain, recombinants were obtained in experiments using identical *oxa* genes, whereas in the 2-AP treated strain, recombination also happened
20 between 5% divergent genes.

Table 4: In vivo recombination frequencies obtained between oxa genes in both wild-type and 2-AP treated *E. coli* strains.

| Genetic divergence (%) | Recombination frequencies | |
|---------------------------|---------------------------|---------------------|
| | wt strain | 2-AP treated strain |
| 0 | 10^{-4} | 10^{-5} |
| 5 | --- | 10^{-6} |

It was noteworthy that recombination occurred between 22% divergent genes, in which the longest stretch of sequence identity was 22 nucleotides (oxa5/oxa11) and 18 nucleotides (oxa5/oxa7), respectively (see figures 5 and 6). It has been reported that recombination becomes inefficient below a minimal length of homology, known as MEPS (minimal efficient processing segment). Length of MEPS varies dependent on the recombination pathway but it has been described to range from 23 to 90 base pairs (5).

Sequence analyses of genes R1 and R2 carried by 54 hybrid plasmids obtained in experiments involving either 5% or 22% divergent genes, showed that 46 of those hybrid plasmids were different to each other. This result indicates that they corresponded to different events of recombination, and consequently, that a high degree of genetic diversity was generated by in vivo recombination.

Most of recombinant genes were generated by non-reciprocal single crossovers in different stretches of sequence identity ranging from 4

to 101 nucleotides. In some cases, multiple crossovers were observed, producing either R1 or R2 mosaic genes. Recombinant genes obtained between *oxa7/oxa5* and *oxa11/oxa5* are depicted in figure 4.

- 5 The comparison of DNA and deduced amino acid sequences of recombinant genes revealed as well that 53% of them corresponded to new *oxa* genes (see table 4). Since no frameshifts or stop codons were generated during recombination, they could putatively encode for 38 new functional β -lactamases.
- 10 Table 4: Comparison of nucleotide and deduced amino acid sequences of recombinant *oxa* genes obtained by *in vivo* recombination

| Sequence comparisons | New genes | New proteins |
|----------------------|--------------|--------------|
| R1 | 33/54 | 21/54 |
| R2 | 25/54 | 17/54 |
| Total | 58/108 (53%) | 38/108 (35%) |

Example 2

In vivo recombination of heterologous genes a double plasmid system (Escherichia coli/Bacillus subtilis) whereby recombinant DNA sequences are selected by resistance to tetracycline

- 5 Another double plasmid system was designed to verify the results obtained in example 1 and facilitate cloning and recombination of genes. This system is based on the selection of cells resistant to tetracycline.

1. Bacterial strains and plasmids

- 10 Bacterial strains and plasmids used in this example are shown in table 5 and table 6, respectively.

Table 5: Bacterial strains

| Strains | Genotype | Reference or source |
|--|--|--|
| <i>E. coli</i> AB1157 Nal^R hsd^- | $\text{thr1 leu6 proA2 his4,}$ thil argE3 lacY1 galK2 ara14 xyl15 $\text{mtl1 tsx33 str31 su-}$ $\text{pE44thr}^+ \text{hsdR}^- \text{B nal}^R$ | M. Radman strain collection |
| <i>E. coli</i> MIXP1 | As AB1157 $\text{Nal}^R R^-$ but $\text{mutS::Tn5 (kan}^R)$ | mutS allele from M. Radman strain collection |
| <i>E. coli</i> DH10B | $F^- mcrA \Delta(mrr$ $\text{hsdRMS-mcrBC) } \phi 80$ $\text{lacZ}\Delta\text{m15 } \Delta\text{lacX74}$ $\text{recA1 endA1 } \alpha-$ $\text{raD139 } \Delta(\text{ara,}$ $\text{leu})7697 \text{ galU galK } \lambda^-$ rpsL nupG | Invitrogen Life Technologies Cat 18290-015 |
| <i>B. subtilis</i> 1A423 | $\text{argGH15 leuB8 recA4}$ $\text{thr-5 hsd}_{RI} RM$ | Bacillus Genetic stock center (Ohio State University, USA) |

Table 6: Plasmids

| Plasmids | Selection |
|--|------------------------------------|
| <i>E. coli</i> plasmid pMIX100 | Cm ^R |
| <i>B. subtilis</i> plasmid pMIX101 | Tc ^R , Erm ^R |
| transformation efficiency control plasmids pMIX 102 and pMIX 103 | Amp ^R , Tc ^R |

2. Results

2.1 Construction of the *E. coli* plasmid pMIX100

- 5 The plasmid pMIX100 carries the origin of replication of pACYC184, as well as its chloramphenicol-resistance gene. pMIX100 carries also the gene *lacZ* from pBluescript SK+, which facilitates selection in cloning experiments. *lacZ* encodes for a fragment of the β-galactosidase providing α-complementation for blue/white selection
- 10 of recombinants in a medium containing X-Gal and IPTG. Thus, colonies harboring pMIX100 should be blue in this medium, those carrying plasmid with a gene inserted in the polylinker (MCS) should be white. The physical map of plasmid pMIX100 is shown in figure 5.

2.2 Construction of the *B. subtilis* plasmid pMIX101

- 15 Plasmid pMIX101 is a derivative of the *B. subtilis* plasmid pIL253. Since the pIL253 marker *Erm*^R, which confers resistance to erythromycin is not useful in *E. coli* the tetracycline-resistance marker al-

lowing the selection of hybrid molecules in recombination experiments was introduced. The tetracycline-resistance gene was amplified from the plasmid pACYC184. Thus, pMIX101 carries two markers: *Erm*^R, as a selection marker in *B. subtilis* for cloning of target genes, and *Tc*^R, to select recombinant hybrid molecules in *E. coli*.
5 The physical map of plasmid pMIX101 is shown in figure 6.

2.3 Construction of the transformation efficiency control plasmids pMIX102 and pMIX103

Since the *B. subtilis* vector is not able to replicate in *E. coli*, a transformation efficiency control is necessary to estimate recombination frequency. Since recombinants will be selected by tetracycline resistance, the same marker must be present in the control vector. Plasmid pMIX102, a derivative of pBluescript SK+ contains the *Tc*^R gene amplified from pACYC184. In pMIX102 the *Tc*^R gene is driven by the
10 *plac* promoter. The physical map of plasmid pMIX102 is shown in figure 7.
15

In the second control vector pMIX103 the *Tc*^R gene is cloned in the opposite direction than *plac*. Therefore this gene is expressed from its own promoter. The physical map of plasmid pMIX103 is shown in
20 figure 8.

2.4 Cloning of oxa7, oxa11 and oxa5 into the *E. coli* plasmid pMIX100

To verify the results of the recombination experiments between 0%,
25 5% and 22% divergent-oxa genes obtained in example 1 oxa7,
oxa11 and oxa5 were cloned into the *E. coli* plasmid pMIX100.
Those genes were amplified using primers containing either *Pst*I or

Xhol at their 5' ends. After digestion with those enzymes, the amplified DNA fragments were independently ligated with pMIX100, which was previously cut with *Pst*I + *Xhol*. Competent cells of *E. coli* DHB10 were electroporated with the ligation mixtures, and colonies 5 harboring positive clones were phenotypically selected by chloramphenicol-resistance/ white color on LB plates containing Cm (30 µg/ml) + X-Gal (80 µg/ml) + IPTG (0.5 mM). For each oxa-cloning, five of such transformants were analyzed.

Plasmid DNAs were obtained and analysed by restriction mapping. 10 Results confirmed that pMIX104 carries oxa7, pMIX106 carries oxa11 and pMIX107 carries oxa5.

For cloning into the *B. subtilis* plasmid pMIX101 oxa7 was obtained as a 0.9-kb fragment from pMIX104 by restriction with *Pst*I and *Xhol* and ligated with pMIX101, which was previously cut with the same 15 enzymes. After transformation of *B. subtilis* 1A423 competent cells with the ligation product, cells were selected in LB containing 0.5 µg/ml of erythromycin (Erm). Plasmid DNA was obtained from 24 transformants and analysed by restriction. Results confirmed that all clones contained the 7-kb plasmid pMIX105.

20 The strategy for cloning the oxa7, oxa11 and oxa5 genes into the *E. coli* plasmid pMIX100 and into the *B. subtilis* plasmid pMIX101 is shown in figure 9.

2.5 In vivo recombination of oxa genes in wt compared to MutS⁻ mutant *E. coli*

25 Competent cells of *E. coli* AB1157 *hsdR*⁺ and its MMR⁺ mutant, *E. coli* AB1157 *hsdR*⁺ *CΔmutS*, were independently transformed by e-

lectroporation with plasmids pMIX104 (*oxa7*), pMIX106 (*oxa11*) and pMIX107 (*oxa5*).

- To recombine *oxa* genes, *E. coli* strains harboring those plasmids were electroporated with pMIX105 (*oxa7*). The *B. subtilis* plasmid 5 pMIX101 was used as a negative control, whereas the *E. coli* vector pMIX102, carrying Tc^R marker, was used as the control of transformation efficiency under tetracycline selection. All DNAs were UV-irradiated prior to electroporation in order to increase recombination frequency.
- 10 Recombination frequencies were calculated by dividing the transformation efficiency obtained with *B. subtilis* plasmids per the transformation efficiency obtained with control vector pMIX92. Transformation efficiencies were in turn calculated as number of colony forming units (cfu) obtained per µg of DNA on previously described conditions.
- 15

Results obtained are summarized in table 7. In the wild-type strain, recombinants were obtained in experiments using either identical or 5% divergent *oxa* genes, whereas in the MutS⁻ mutant, recombination also happened between 22% divergent genes.

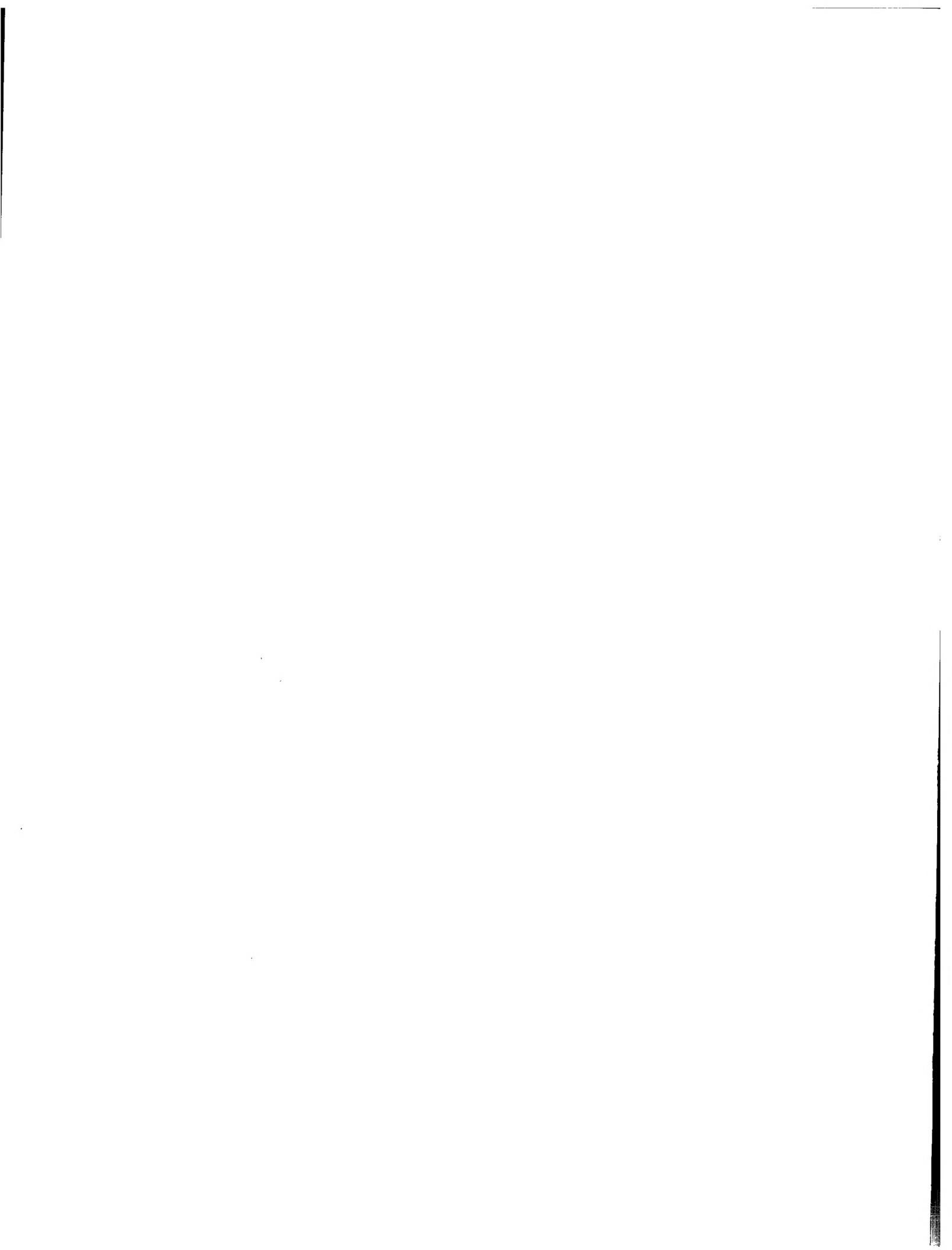
Table 7: In vivo recombination frequencies obtained between oxa genes in both wild-type and MutS⁻ *E. coli* strains.

| Genetic divergence (%) | Recombination frequencies | |
|---------------------------|------------------------------------|------------------------------------|
| | wt strain | MutS ⁻ strain |
| 0 | 10 ⁻⁵ -10 ⁻⁶ | 10 ⁻⁶ |
| 5 | --- | 10 ⁻⁵ -10 ⁻⁶ |
| 22 | --- | --- |

Recombination frequencies were consistent with those obtained with
5 the former double plasmid system.

To confirm recombination, pMIX105 transformants were grown in liquid media containing tetracycline (12,5 µg/ml). Restriction analyses of plasmid DNA obtained from these cultures showed that they harbored dimers carrying oxa-recombinant genes, together with resident plasmids (either pMIX104 or pMIX106). Moreover, recombinant genes R1 and R2 were successfully amplified by PCR from both colonies and plasmid DNA using specific primers.
10

Results demonstrate that recombinants were generated with a second double plasmid system using tetracycline resistance as selective pressure.
15



Claims

1. Process for generating and detecting recombinant DNA sequences in prokaryotes comprising the steps of:
 - a) generating a first prokaryotic cell containing a recipient DNA molecule, which comprises a first DNA sequence to be recombined and which can autonomously replicate in the prokaryotic cell, and a donor DNA molecule, which comprises a second DNA sequence to be recombined and at least a first marker sequence encoding a gene product and which cannot autonomously replicate in the prokaryotic cell,
5
10
 - b) cultivating the first prokaryotic cell under selective conditions which only allow the growth and/or propagation of the cell if the gene product of the first marker sequence is expressed, and
 - c) isolating a second prokaryotic cell grown and/or propagated under selective conditions and containing a hybrid DNA molecule with the at least first marker sequence and a first and a second recombined DNA sequences due to recombination between the first and the second DNA sequences.
15
- 20 2. Process according to claim 1, wherein the donor DNA molecule and the recipient DNA molecules are different linear or circular DNA structures, in particular different plasmids or bacteriophages.
3. Process according to claim 1 or 2, wherein the recipient DNA molecule is an plasmid, which can replicate in *Escherichia coli*.

4. Process according to claim 3, wherein the recipient DNA molecule is the *E. coli* plasmid pACYC184 or the *E. coli* plasmid pMIX100 or a derivative thereof.
5. Process according to any one of claims 1 to 4, wherein the donor DNA molecule does not have an origin of replication.
6. Process according to any one of claims 1 to 4, wherein the donor DNA molecule has a non-functional origin of replication.
7. Process according to claim 6, wherein the donor DNA molecule and/or its origin of replication are derived from a prokaryotic species other than the prokaryotic species, in cells of which the donor DNA molecule is introduced.
8. Process according to claim 7, wherein the donor DNA molecule is a *Bacillus subtilis* plasmid, which cannot replicate in *E. coli*.
9. Process according to claim 7 or 8, wherein the donor DNA molecule is the *B. subtilis* plasmid pMIX91 comprising the *spec^R* marker and the *phleo^R* marker or the *B. subtilis* plasmid pMIX101 comprising the *tc^R* marker.
10. Process according to claim 6, wherein the function of the origin of replication of the donor DNA is impaired by a mutation.
11. Process according to any one of claims 1 to 10, wherein the first marker sequence of the donor DNA structure is selected from the group consisting of a nutritional marker, an antibiotic resistance marker and a sequence encoding a subunit of an enzyme.

- 12: Process according to claim 11, wherein the gene product of the first marker sequence confers resistance to an antibiotic to a cell which is sensitive to that antibiotic.
13. Process according to claim 11 or 12, wherein the first marker sequence is *spec^R* the gene product of which confers to a cell resistance to spectinomycin or *phleo^R* the gene product of which confers to a cell resistance to phleomycin or *tc^R* the gene product of which confers to a cell resistance to tetracycline.
14. Process according to any one of claims 1 to 13, wherein the donor DNA molecule contains a second marker sequence.
15. Process according to any one of claims 1 to 14, wherein the recipient DNA molecule contains a third marker sequence and optionally a fourth marker sequence.
16. Process according to claim 14 or 15, wherein the second, third and fourth marker sequences are protein-coding or non-coding sequences selected from the group consisting of nutritional markers, pigment markers, antibiotic resistance markers, antibiotic sensitivity markers, restriction enzymes sites, primer recognition sites and sequences encoding a subunit of an enzyme.
17. Process according to claim 16, wherein the gene products of the third and fourth marker sequences of the recipient DNA molecule confer resistance to an antibiotic to a cell which is sensitive to that antibiotic.
18. Process according to claim 17, wherein the gene product of the third marker sequence confers to a cell resistance to tetracycline.

19. Process according to claim 17, wherein the gene product of the fourth marker sequence confers to a cell resistance to chloramphenicol.
 20. Process according to any one of claims 1 to 19, wherein the first
5 and the second DNA sequences to be recombined diverge by at least two nucleotides.
 21. Process according to any one of claims 1 to 20, wherein the first and the second DNA sequences to be recombined are naturally occurring sequences.
- 10 22. Process according to claim 21, wherein the first and/or the second DNA sequences to be recombined are derived from viruses, bacteria, plants, animals and/or human beings.
- 15 23. Process according to any one of claims 1 to 20, wherein the first and/or the second DNA sequences to be recombined are artificial sequences.
24. Process according to any one of claims 1 to 23, wherein each of the first and the second DNA sequences to be recombined comprises one or more protein-coding sequences and/or one or more non-coding sequences.
- 20 25. Process according to any one of claims 1 to 24, wherein the first prokaryotic cell is generated by simultaneously or sequentially introducing the recipient DNA molecule and the donor DNA molecule into a prokaryotic cell.

26. Process according to claim 25, wherein the recipient and donor DNA molecules are introduced into the prokaryotic cell via transformation, conjugation, transduction, sexduction and/or electroporation.
27. Process according to any one of claims 1 to 26, wherein the first 5 prokaryotic cell is cultivated in the presence of at least one antibiotic to which the gene product of the first marker sequence confers resistance.
28. Process according to claim 27, wherein the first prokaryotic cell is additionally cultivated in the presence of a second, a third and/or a 10 fourth antibiotics to which the gene products of the second marker sequence, the third marker and the fourth marker sequence, respectively, confer resistance.
29. Process according to any one of claims 1 to 28, wherein the prokaryotic cell is a cell of an archaebacterium or an eubacterium.
- 15 30. Process according to claim 29, wherein the eubacterium is a gram-negative bacterium, a gram-positive bacterium or an cyanobacterium.
31. Process according to claim 30, wherein the gram-negative bacterium is *Escherichia coli*.
- 20 32. Process according to any one of claims 1 to 31, wherein the prokaryotic cell has a functional mismatch repair system.
33. Process according to any one of claims 1 to 31, wherein the prokaryotic cell is transiently or permanently deficient in the mismatch repair system.

34. Process according to claim 33, wherein the transient or permanent deficiency of the mismatch repair system is due to a mutation, a deletion, and/or an inducible expression or repression of one or more genes involved in the mismatch repair system, a treatment with an agent that saturates the mismatch repair system and/or a treatment with an agent that globally knocks out the mismatch repair.
- 5
35. Process according to claim 33 or 34, wherein the prokaryotic cell has a mutated *mutS* gene and/or mutated *mutL* gene.
36. Process according to any of claims 1 to 35, wherein the first and 10 the second recombined DNA sequences contained in the hybrid DNA molecule of the second prokaryotic cell are selected and/or isolated and/or analysed.
- 15
37. Process according to claim 36, wherein the first and the second recombined DNA sequences are isolated by restriction enzyme cleavage.
38. Process according to claim 36, wherein the first and the second recombined DNA sequences are amplified by PCR.
39. Process according to any one of claims 36 to 38, wherein the isolated first and second recombined DNA sequences are inserted 20 into a donor DNA molecule and a recipient DNA molecule, respectively, and subjected another round of recombination.
40. *Bacillus subtilis* plasmid pMIX91 which comprises the *spec^R* marker and the *phleo^R* marker and the restriction sites *Scal*, *PpuMI* and *EcoO109I* for inserting a foreign DNA sequence.

41. *Bacillus subtilis* plasmid pMIX101 which comprises the *tc^R* marker sequence and the restriction sites *Xba*I and *Pst*I for inserting a foreign DNA sequence.
42. Use of the *B. subtilis* plasmids pMIX91 or pMIX101 as donor 5 DNA molecules in a process according to any one of claims 1 to 40 for generating and/or detecting recombinant DNA sequences in a prokaryotic host cell, preferably in an *E. coli* cell.
43. Use of the *E. coli* plasmids pACYC184 or pMIX100 or a derivative thereof as recipient DNA molecule in a process according to any 10 one of claims 1 to 40 for generating and/or detecting recombinant DNA sequences in a prokaryotic host cell, preferably in an *E. coli* cell.
44. Kit comprising at least a first container which comprises cells of the *E. coli* strain AB1157 or the *E. coli* strain MXP1 or the *E. coli* 15 strain DHB10, a second container which comprises cells of the *E. coli* strain AB1157 containing plasmid pACYC184 or cells of the *E. coli* strain DHB10 containing plasmid pMIX100 and a third container comprising cells of the *B. subtilis* strain DSM4393 containing plasmid pMIX91 or cells of the *B. subtilis* strain 1A423 containing plasmid 20 pMIX101.
45. Kit comprising at least a first container which comprises cells of the *E. coli* strain AB1157 or the *E. coli* strain MXP1 or the *E. coli* strain DHB10, a second container comprising DNA of plasmid pA-CYC184 or plasmid pMIX100 and a third container comprising DNA 25 of plasmid pMIX91 or plasmid pMIX101.

46. A process for producing a hybrid gene and/or a protein encoded by a hybrid gene in a prokaryotic cell, wherein a process according to any one of claims 1 to 39 is carried out and the hybrid gene and/or the protein encoded by the hybrid gene is produced in the prokaryotic cell and the hybrid gene and/or the encoded protein is selected in the prokaryotic cell and/or isolated therefrom after expression.
- 5
47. Hybrid gene obtainable by a process according to claim 46.
48. Protein, which is encoded by a hybrid gene according to claim 47 and which is obtainable by a process according to claim 46.

Abstract

**Generation of recombinant genes in prokaryotic cells by using two
extrachromosomal elements**

The present invention relates in general to methods for generating and detecting recombinant DNA sequences in prokaryotic cells, in particular bacteria, by using two different extrachromosomal elements and extrachromosomal elements, in particular plasmids that can be used for conducting the inventive methods. DNA sequences for which these methods are relevant include protein-encoding and non-coding sequences.

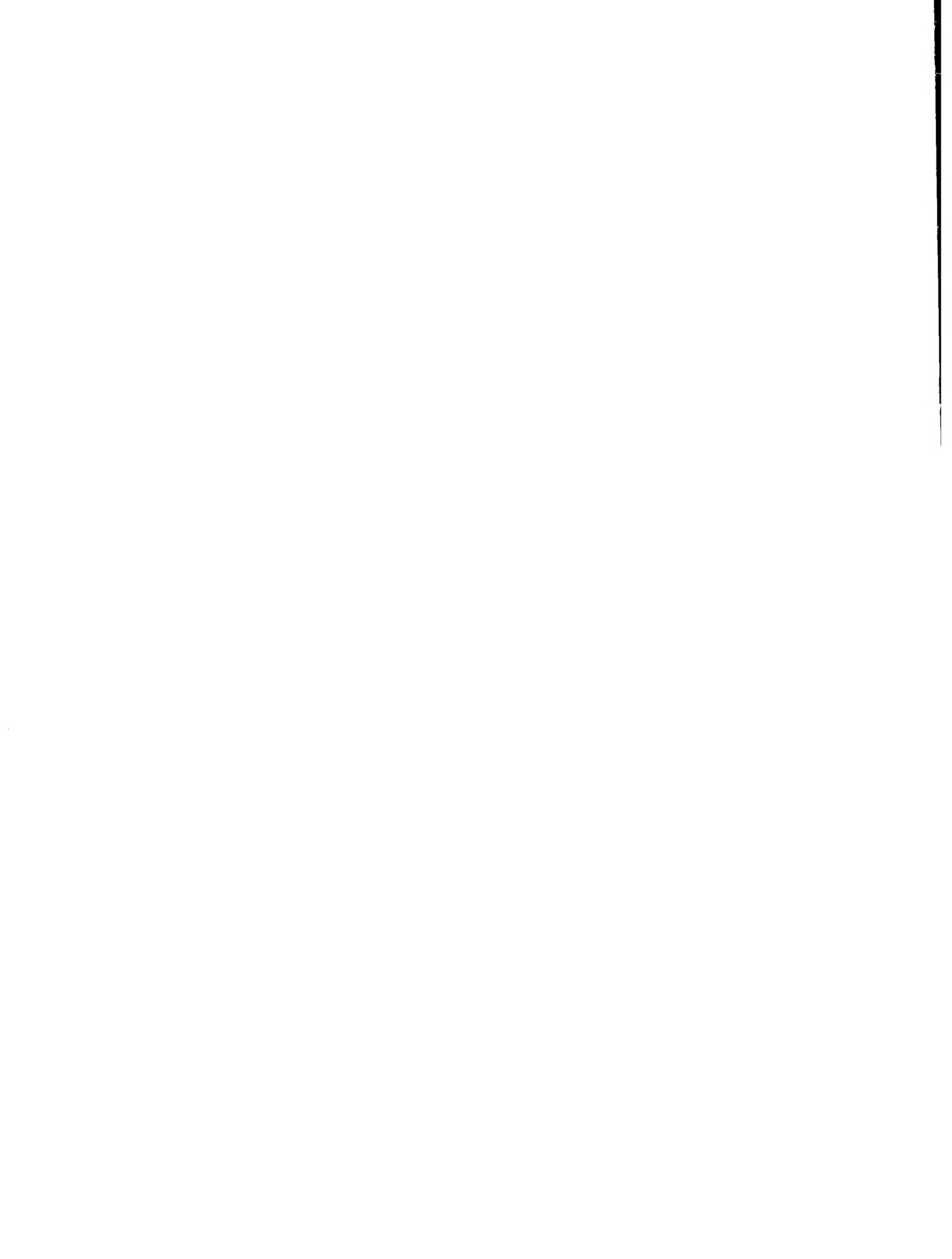
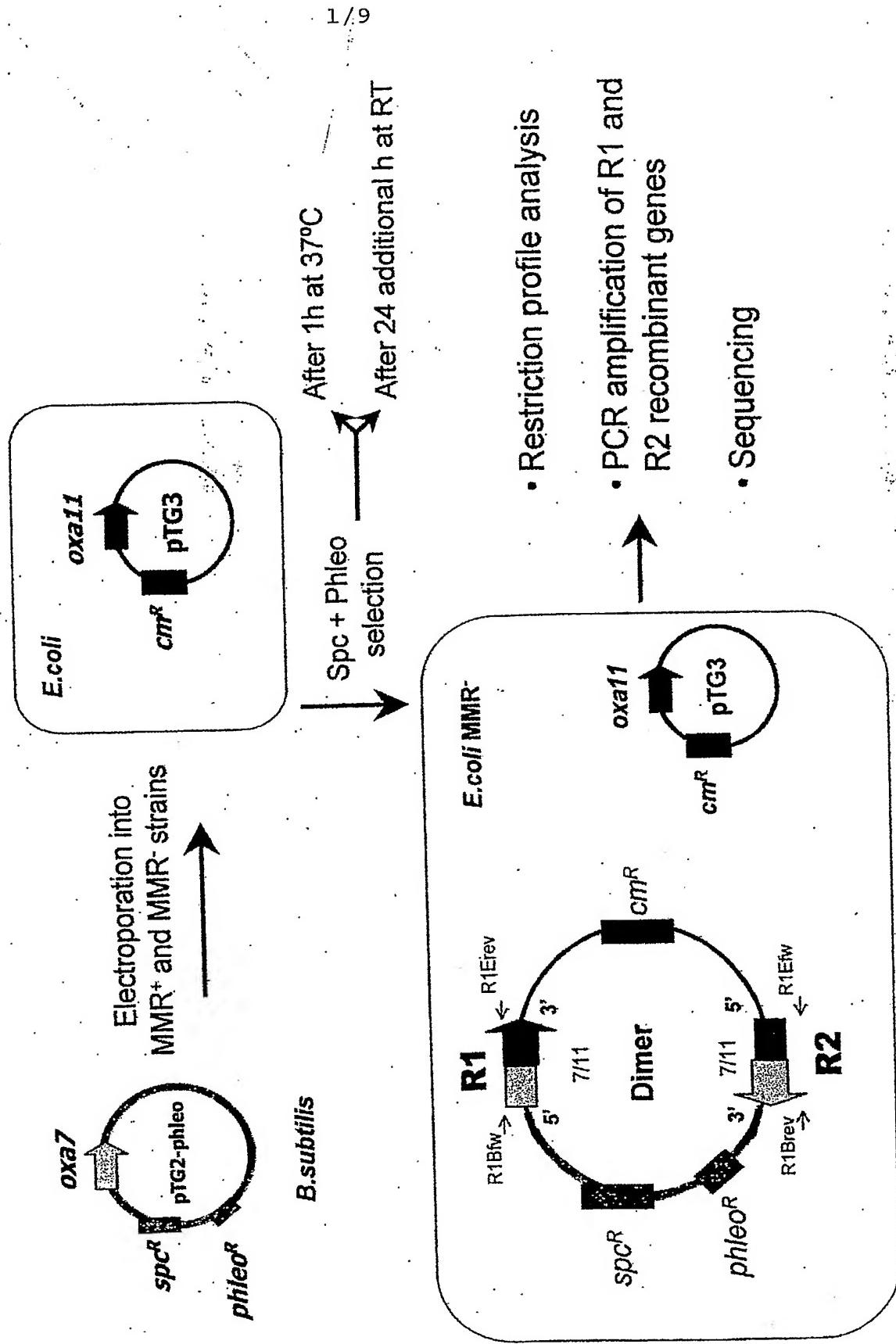


Figure 1



Recombination frequency in MMR- strain $\approx 10^{-6}$

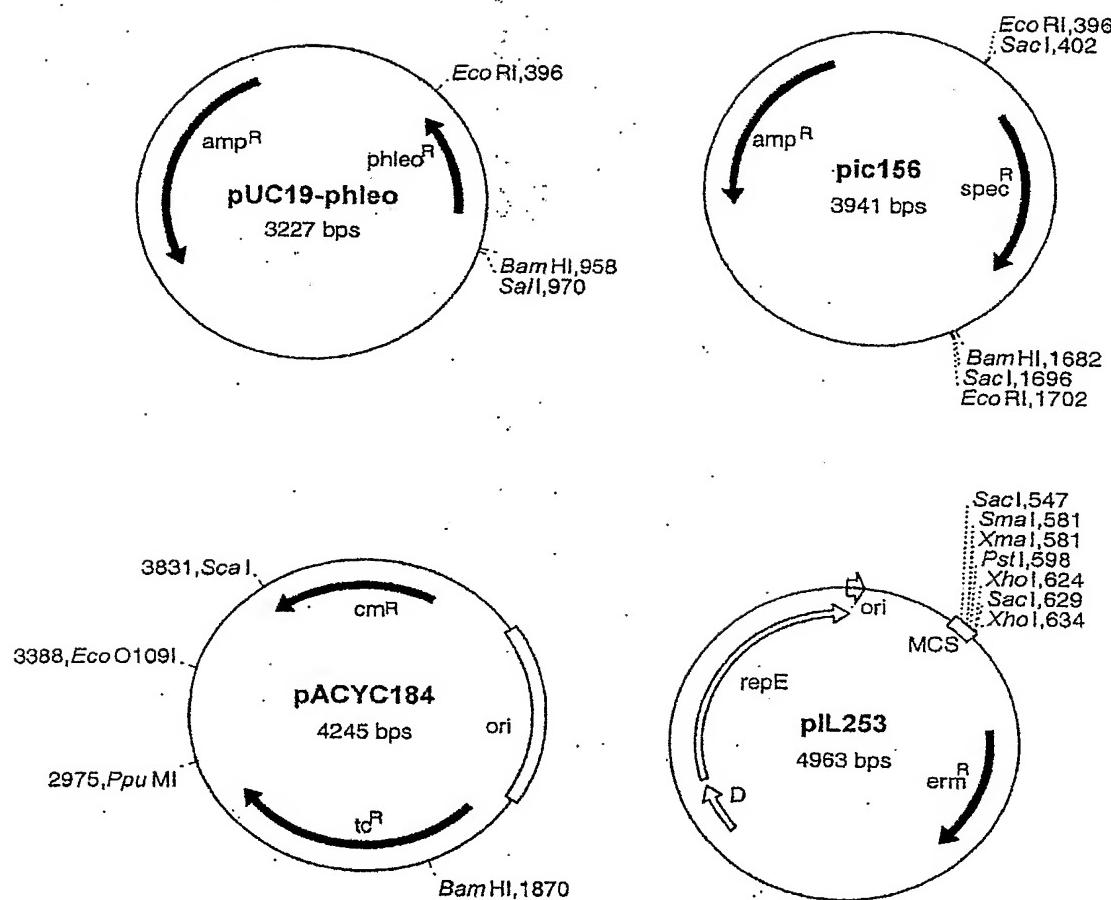
Figure 2

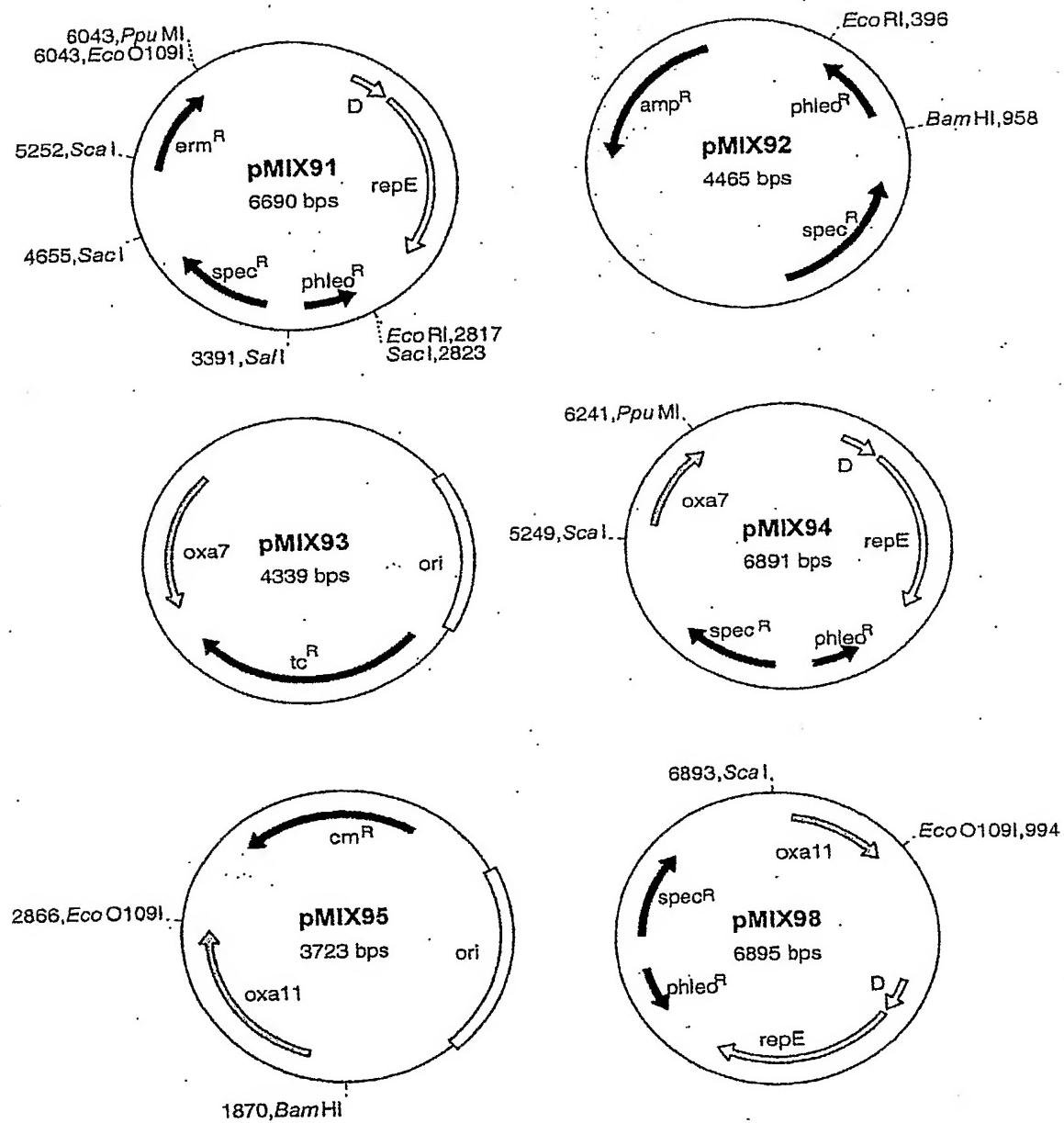
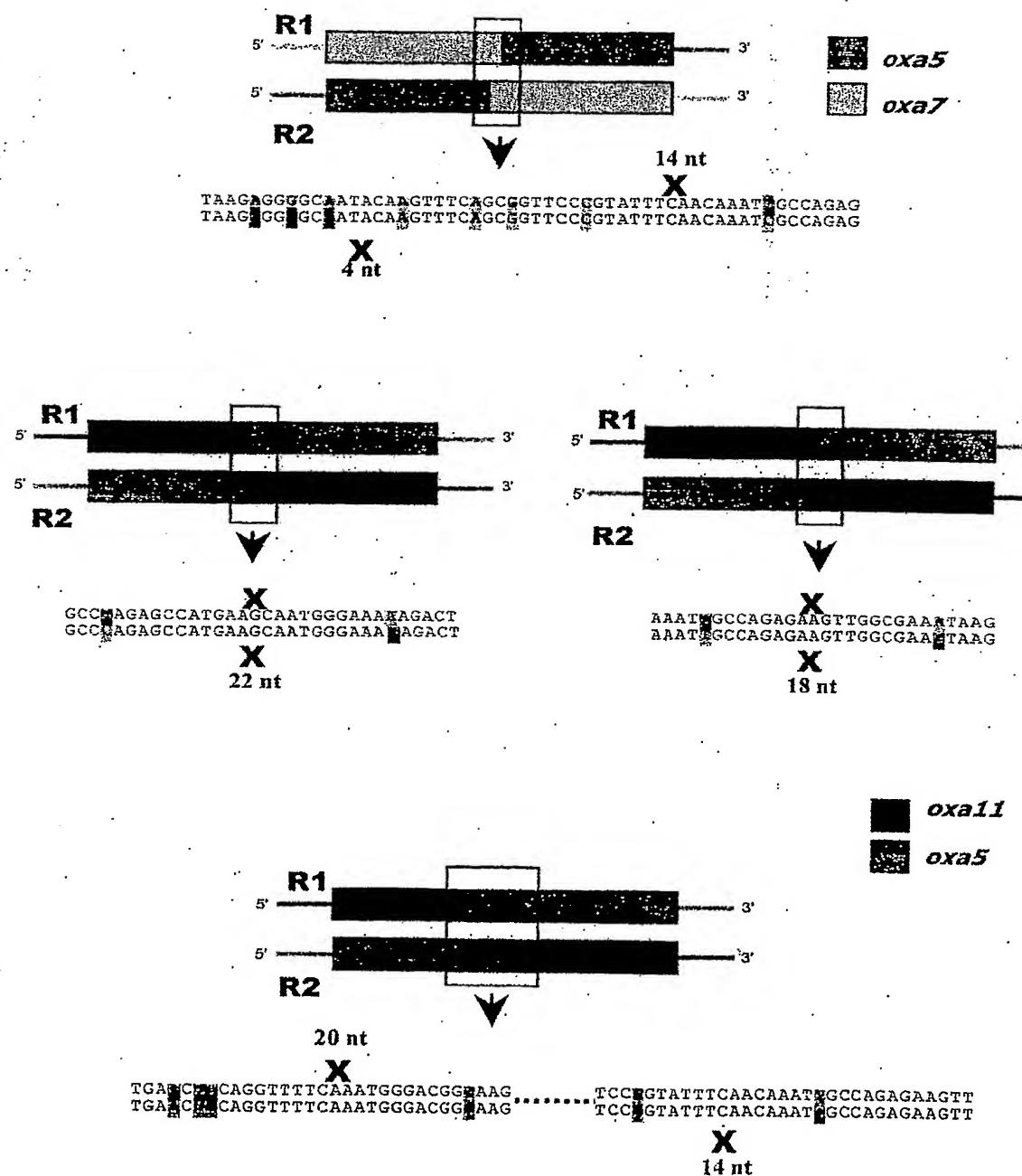
Figure 3

Figure 4

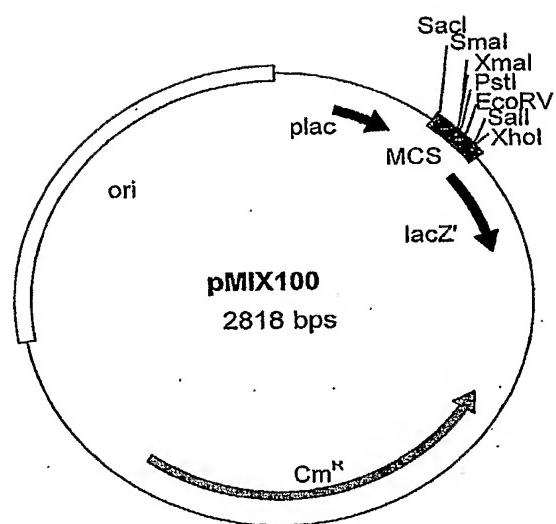


Figure 5

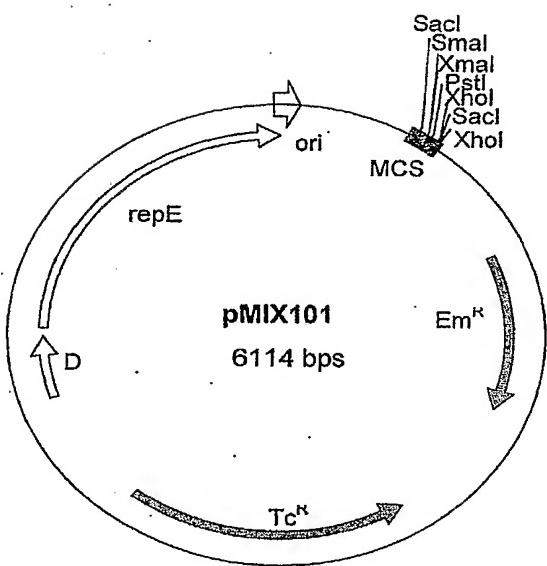


Figure 6

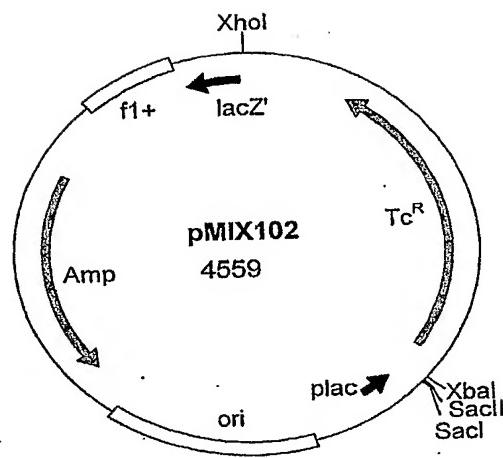


Figure 7

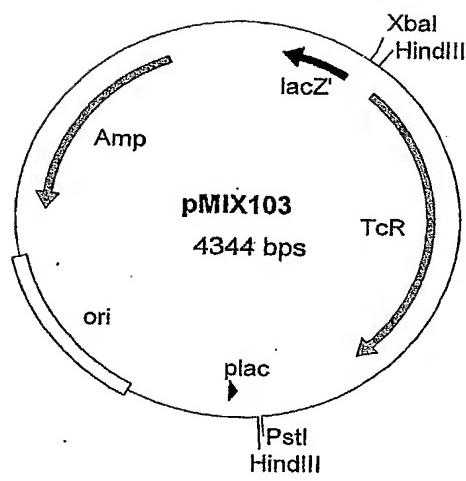
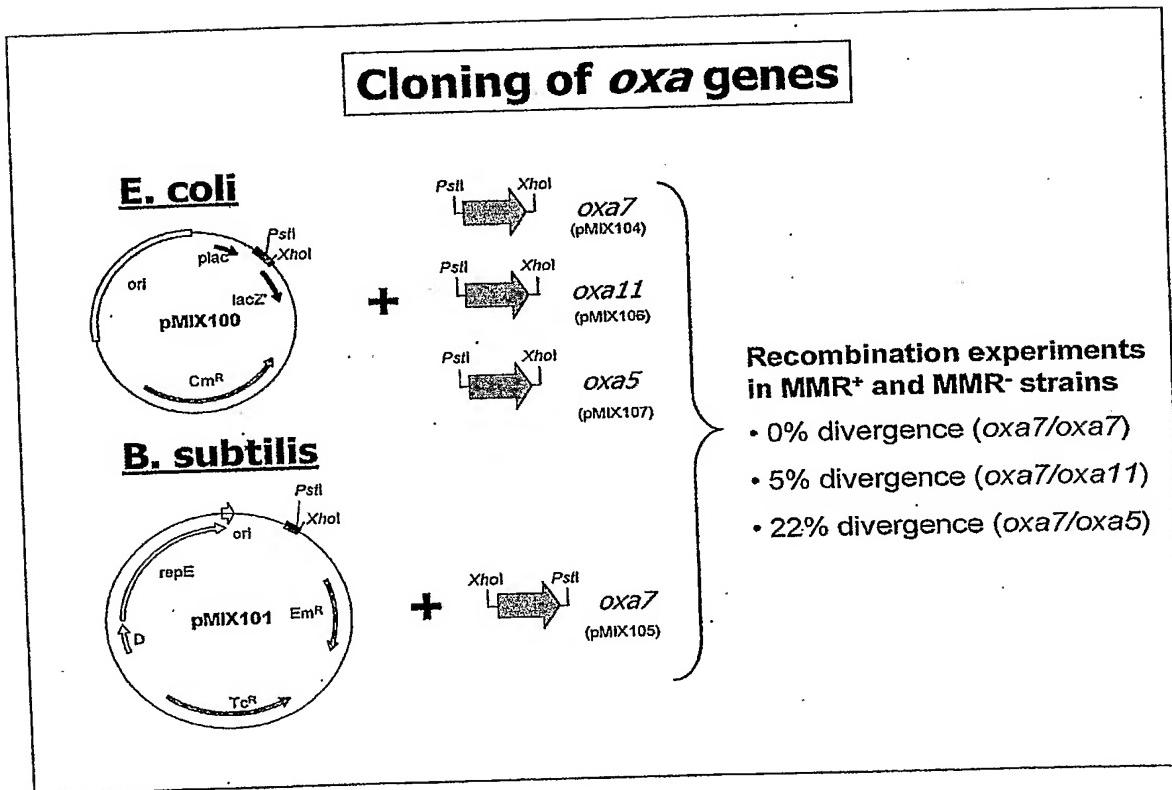
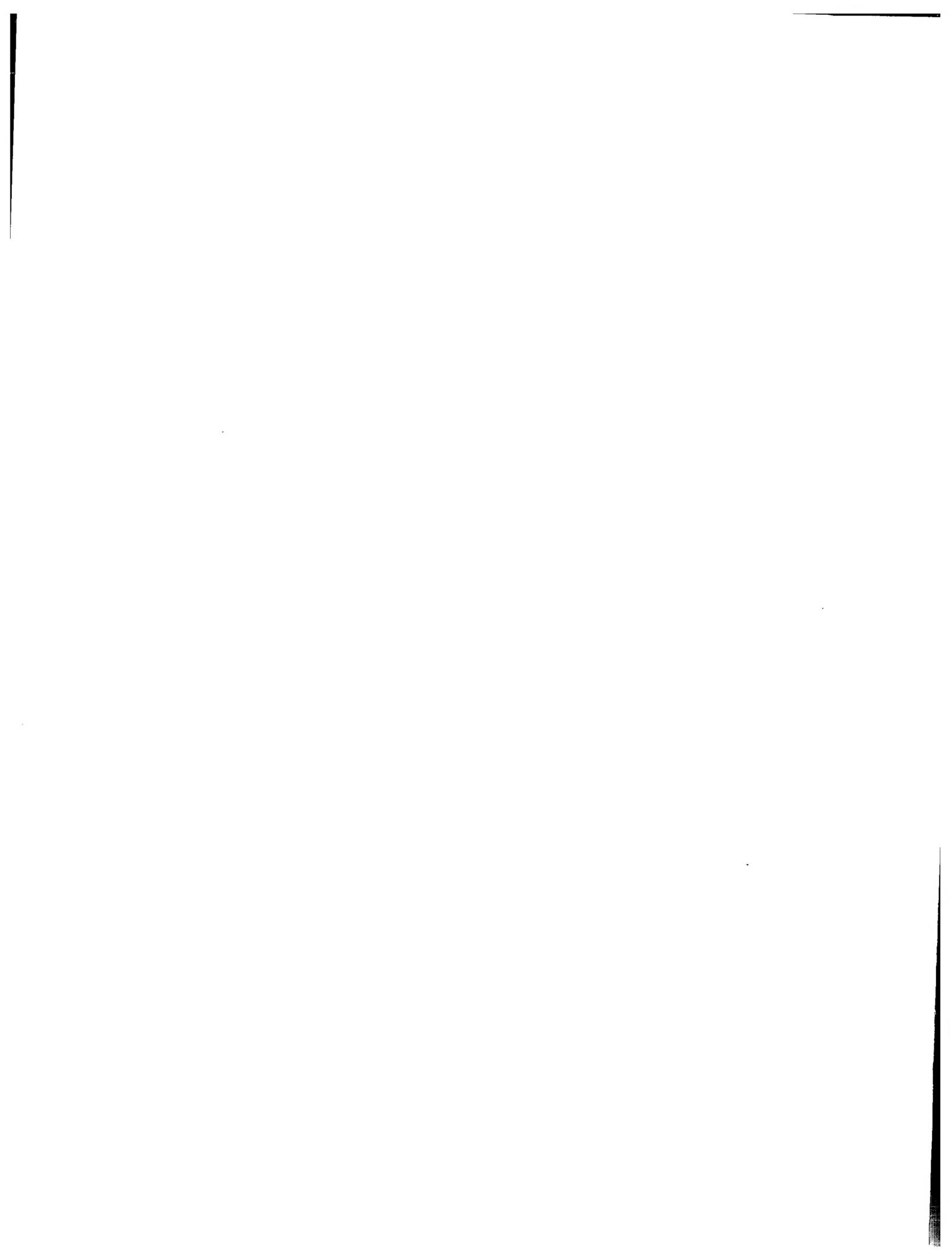


Figure 8

**Figure 9**



SEQUENCE LISTING

<110> Mixis France S.A.

<120> Generation of recombinant genes in prokaryotic cells by
using two extrachromosomal elements

<130> 18249

<140>

<141>

<160> 17

<170> PatentIn Ver. 2.1

<210> 1

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 1

gcmcagttact cctcactcg ggccggaaaag g

31

<210> 2

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 2

gcgcagggtcc cgtttgagct caggccgcg

29

<210> 3

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 3

gcggatccat gcctgcaggg acgcctttg

29

<210> 4

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 4

gcgcagggcc tggtcacgat gctgtacttt gtg

33

<210> 5
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 5
gcggatcctg ttagccacca aggtacca

28

<210> 6
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 6
gcmcaggccc tttagccacc aatgatgatg c

31

<210> 7
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 7
gcggatcccc cttaacaaaa ccaatac

27

<210> 8
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 8
gcmcaggccc taagggttgg gcgattttg.

29

<210> 9
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 9
gcmcagtact atgcctgcag ggacgcctt g

31

<210> 10
<211> 33

| | | |
|--|--|----|
| <212> DNA | | |
| <213> Artificial Sequence | | |
| | | |
| <220> | | |
| <223> Description of Artificial Sequence: Primer | | |
| | | |
| <400> 10 | | 33 |
| tcgcgggacc tggcacgat gctgtacttt gtg | | |
| | | |
| <210> 11 | | |
| <211> 30 | | |
| <212> DNA | | |
| <213> Artificial Sequence | | |
| | | |
| <220> | | |
| <223> Description of Artificial Sequence: Primer | | |
| | | |
| <400> 11 | | 30 |
| gcgcagtact gggcgaaccc ggagcctcat | | |
| | | |
| <210> 12 | | |
| <211> 22 | | |
| <212> DNA | | |
| <213> Artificial Sequence | | |
| | | |
| <220> | | |
| <223> Description of Artificial Sequence: Primer | | |
| | | |
| <400> 12 | | 22 |
| aagaaggagt gattacatga ac | | |
| | | |
| <210> 13 | | |
| <211> 19 | | |
| <212> DNA | | |
| <213> Artificial Sequence | | |
| | | |
| <220> | | |
| <223> Description of Artificial Sequence: Primer | | |
| | | |
| <400> 13 | | 19 |
| cgcatctcggtcgcagcggttg | | |
| | | |
| <210> 14 | | |
| <211> 20 | | |
| <212> DNA | | |
| <213> Artificial Sequence | | |
| | | |
| <220> | | |
| <223> Description of Artificial Sequence: Primer | | |
| | | |
| <400> 14 | | 20 |
| gcagatccgg aacataatgg | | |
| | | |
| <210> 15 | | |
| <211> 20 | | |
| <212> DNA | | |
| <213> Artificial Sequence | | |

<220>
<223> Description of Artificial Sequence: Primer
<400> 15
tgtcggcaga atgcttaatg

20

<210> 16
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
<400> 16
cactatcgac tacgcgatca

20

<210> 17
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
<400> 17
gcttccccat gataagagc

19